

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 02:04:13 ; Search time 2348 Seconds  
(without alignments)  
10151.277 Million cell updates/sec

Title: US-09-673-918a-1

Perfect score: 819

Sequence: 1 atgcagcttcgaactgcatt.....tgtccaataatccagattct 819

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.mu.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg\_hum.\*

31: em.htg\_inv.\*

32: em.htg\_other.\*

33: em.htg\_mus.\*

34: em.htg\_pln.\*

35: em.htg\_rod.\*

36: em.htg\_mam.\*

37: em.htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	806.4	98.5	831	8	AF352735	AF352735 Forsythia
2	260.8	31.8	933	8	NTAJ3177	AF223177 Nicotiana
3	260.4	31.8	930	8	AF286650	AF286650 Cucumis s
4	260.4	31.8	930	8	AF286651	AF286651 Cucumis s
5	252.6	30.8	1089	8	AF072447	AF072447 Ipomoea t
6	252.6	30.8	1152	8	NTAJ3178	AF223178 Nicotiana
7	250.4	30.6	1064	8	D88121	D88121 Vigna ungu
8	246.2	30.1	1179	8	AF072449	AF072449 Ipomoea t
9	244.4	29.8	3021	8	AB070221	AB070221 Ipomoea t
10	243	29.7	1120	8	AF072448	AF072448 Ipomoea t
11	243	29.7	1144	8	AF072450	AF072450 Ipomoea t
12	223.2	27.3	992	8	AB018559	AB018559 Citrullus
13	206.4	25.2	834	8	AF352734	AF352734 Podophyl
14	201.8	24.6	1176	8	AY099603	AY099603 Arabidops
15	200.8	24.5	75948	8	AC037424	AC037424 Arabidops
16	200	24.4	107109	2	OSJN00137	AL662936 Oryza sat
17	200	24.4	107109	2	OSJN00145	AL662933 Oryza sat
18	188.8	23.1	980	6	EO9193	D29976 Tobacco mRN
19	188.8	23.1	1025	6	EO9193	EO9193 Transcripti
20	187	22.8	111844	2	OSJN00288	AL731639 Oryza sat
21	179.8	22.0	922	8	LES277945	AL7277945 Lycopersi
22	175.8	21.5	2226	8	AB070220	AB070220 Ipomoea t
23	170.8	20.9	79867	8	ATT18N14	AL132968 Arabidops
24	170.6	20.8	884	8	LEU21801	U21801 Lycopersico
25	164	20.0	90864	8	AF004949	AF004949 Lotus jap
26	162.8	19.9	944	8	AF097651	AF097651 Pisum sat
27	159.6	19.5	969	8	AF053638	AF053638 Pisum sat
28	156.4	19.1	780	8	AJ345026	AJ345026 Digitalis
29	155	18.9	1050	8	AF370319	AF370319 Arabidops
30	153.4	18.7	106329	8	AC004411	AC004411 Arabidops
31	153.4	18.7	106329	8	AC004411	AC004411 Arabidops
32	150.2	18.3	774	6	AX412402	AX412402 Sequence
33	150.2	18.3	774	6	AX412739	AX412739 Sequence
34	150	18.3	77401	8	AB026657	AB026657 Arabidops
35	147.8	18.0	143407	8	AF003825	AF003825 Oryza sat
36	135.8	16.6	817	8	AF053639	AF053639 Pisum sat
37	130.8	16.0	999	8	AY052216	AY052216 Arabidops
38	127.8	15.6	911	8	SLU53828	U53828 Silene lati
39	126	15.4	151914	2	AP004333	AP004333 Oryza sat
40	119.6	14.6	929	8	SLU53829	U53829 Silene lati
41	118.8	14.5	354	8	AF188188	AF188188 Cucumis s
42	118	14.4	1124	8	SLU53827	U53827 Silene lati
43	114.8	14.0	1187	6	AR007559	AR007559 Sequence
44	113.6	13.9	153308	8	AC118668	AC118668 Genomic s
45	109.4	13.4	163154	2	AC092557	AC092557 Oryza sat

ALIGNMENTS

RESULT 1  
AF352735

LOCUS  
DEFINITION

AF352735

AF352735

AF352735

AF352735

AF352735

AF352735

AF352735

AF352735

AF352735

AF352735

AF352735

AF352735

AF352735

AF352735

AF352735

AF352735

831 bp mRNA linear PLN 23-APR-2001  
Forsythia x intermedia stem secoisolariciresinol dehydrogenase  
mRNA, partial cds.  
GI:13752457  
Forsythia x intermedia.  
Forsythia x intermedia  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Lamiales; Oleaceae; Forsythia.  
1 (bases 1 to 831)  
Xia, Z.Q., Costa, M.A., Pelissier, H.C., Davin, L.B. and Lewis, N.G.

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
99164-6340, USA

Secoisolaricresinol dehydrogenase purification, cloning, and functional expression. Implications for human health protection J. Biol. Chem. 276 (16), 12614-12623 (2001)

2 (bases 1 to 831)  
Xia, Z.Q., Costa, M.A., Davin, L.B. and Lewis, N.G.  
Direct Submission  
Submitted (26-FEB-2001) Institute of Biological Chemistry, Washington State University, 461 Clark Hall, Pullman, WA

FEATURES  
Source  
1..831  
/organism="Forsythia x intermedia"  
/db\_xref="taxon:55183"  
1..>831  
/note="SDH-F1321; NAD-dependent; catalyzes enantiospecific conversion of (-)-secoisolaricresinol into (-)-matairesinol"  
/codon\_start=1  
/product="stem secoisolaricresinol dehydrogenase"  
/protein\_id="AAK38665.1"  
/db\_xref="GI:13752458"  
/translation="MAATSQVLTATARLEGKVALITGGASGIGETAKLFSQHGAKV AIADVDLGHVSVEAIGTSNVTYIHCDDVTNEDGVKNVADNTVSTYKGLDIFMSNAGI SDNRPRIIDNEKADPERVSVNVTGFLCMKHAARVIMPARSGNIISTASLSNEMGG GSSHAYCGSKHVALRNLAVELGFGIRVNCSPFLGTLALGKFKSGIKNEFEFEN VINFAGNLGKPRFNVEDVANAALYLASDEAKYVSGHNLFDGGFSVCNSVTKVFOYPD S"

BASE COUNT 244 a 169 c 192 g 226 t

ORIGIN

Query Match 98.5%; Score 806.4; DB 8; Length 831;  
Best Local Similarity 99.3%; Pred. No. 4.3e-195;  
Matches 810; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 CAGCTTGAAGTATGCGAAGAGCTGAGAGAAAGTTCCCTTTATACAGAGGA 63  
DB 16 CAGGTTCTACTGCAATCGAAGAGCTGAGAGAAAGTTCCCTTTATACAGAGGA 75  
QY 64 GCCAGTGAATTGAGAACACACAGAAACTCTTCCCAACATGGAGCCAAAGTTGCC 123  
DB 76 GCCAGTGAATTGAGAACACACAGAAACTCTTCCCAACATGGAGCCAAAGTTGCC 135  
QY 124 ATTGCTGATGCCAAGATGAATTAGGTCACTAGTTGTTCGAGGCCATTGGCATTCCAAT 183  
DB 136 ATTGCTGATGCCAAGATGAATTAGGTCACTAGTTGTTCGAGGCCATTGGCATTCCAAT 195  
QY 184 TCCACCTACATCCACTGTGATGTTACTTAATGAAGACGGTGTAAATAATCCCGTGGACAAC 243  
DB 196 TCCACCTACATCCACTGTGATGTTACTTAATGAAGACGGTGTAAATAATCCCGTGGACAAC 255  
QY 244 ACAGTTTCAACCTATGGAACACTGGACATTATGTTTCAGCAATGCAGGAATTTCTGATCCC 303  
DB 256 ACAGTTTCAACCTATGGAACACTGGACATTATGTTTCAGCAATGCAGGAATTTCTGATCCC 315  
QY 304 AACAGGCCCGCATCATAGACAAAGACAGATTTGACGGGTCTCGAGTGTATAT 363  
DB 316 AACAGGCCCGCATCATAGACAAAGACAGATTTGACGGGTCTCGAGTGTATAT 375  
QY 364 GTAACCGAGTTTTCATGATGATGACGACGACGACGATGTTATGATCCACGACGAGT 423  
DB 376 GTAACCGAGTTTTCATGATGATGACGACGACGACGATGTTATGATCCACGACGAGT 435  
QY 424 GGCAACATAATTTCCACTGCTAGTTTAAGCTCAACTATGGGTGGTGTCTTCATATGCC 483  
DB 436 GGCAACATAATTTCCACTGCTAGTTTAAGCTCAACTATGGGTGGTGTCTTCATATGCC 495  
QY 484 TATTGTGGTTCAAAGCATGTTGTTAGCCCTTACTAGGAATCTGGCAGTCGAGTCGGA 543  
DB 496 TATTGTGGTTCAAAGCATGTTGTTAGCCCTTACTAGGAATCTGGCAGTCGAGTCGGA 555  
QY 544 CAATTGTGCATTAGGGTTAATTGTTGTCCTCTTCGCGGCTCCCTAGCGGCTTTAGGCAAG 603

Db 556 CAATTTGGCATTTAGGGTTAATTTGTTGCTCTCCTTCGGGCTTCCTACGGCTTTAGGCAAG 615  
QY 604 AAATTTTCAGGGATTAAAAATGAAGAAGAAATTTGAGAATGTAAATAAACTTTGCGGGAAT 663  
Db 616 AAATTTTCAGGGATTAAAAATGAAGAAGAAATTTGAGAATGTAAATAAACTTTGCGGGAAT 675  
QY 664 TTGAAGGTCGCAAAATTTAATGTTGAGGATGTTGCCAATGCAGCTCTTTATCTGGTAGT 723  
Db 676 CTGAAGGTCGCAAAATTTAATGTTGAGGATGTTGCCAATGCAGCTCTTTATCTGGTAGT 735  
QY 724 GATCAGGCAAAATACGTCGATGACACAAATCTTTCATGATGGAGGCTTCAGGCTCGC 783  
Db 736 GATGAGGCAAAATACGTCGATGACACAAATCTTTCATGATGGAGGCTTCAGGCTCGC 795  
QY 784 AATTCGTATCAAACTGTTTCCAAATTCAGATTCT 819  
Db 796 AATTCGTATCAAACTGTTTCCAAATTCAGATTCT 831

RESULT 2  
LOCUS NTAUJ3177 933 bp mRNA linear PLN 08-FEB-1998  
DEFINITION Nicotiana tabacum mRNA for short chain alcohol dehydrogenase.  
ACCESSION AJ223177  
VERSION AJ223177.1 GI:2739278  
KEYWORDS short chain alcohol dehydrogenase.  
SOURCE Nicotiana tabacum.  
ORGANISM Nicotiana tabacum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
REFERENCE 1 (bases 1 to 933)  
Moenke, G.  
Direct Submission  
Submitted (17-DEC-1997) Moenke G.; Molecular Genetics, Institut fuer Pflanzen-genetik und Kulturpflanzenforschung, Gatersleben, D-06466, GERMANY  
2 (bases 1 to 933)  
Herbers, K., Monke, G., Badur, R. and Sonnwald, U.  
A simplified procedure for the subtractive cDNA cloning of photoassimilate-responsive genes: isolation of cDNAs encoding a new class of pathogenesis-related proteins  
Plant Mol. Biol. 29 (5), 1027-1038 (1995)  
96145513  
PUBMED 855446  
FEATURES  
Location/Qualifiers  
1..933  
/organism="Nicotiana tabacum"  
/variety="SNN"  
/db\_xref="taxon:4097"  
/tissue\_type="leaf"  
5..859  
/codon\_start=1  
/product="short chain alcohol dehydrogenase"  
/protein\_id="CAA11153.1"  
/db\_xref="GI:2739279"  
/translation="MASYSLSPIAKRLKLEKVALITGASGIGAAATLRFVQHGAKVT IADIQNLGTSLVQEIQNEHTIFVHCNVAVESDVQNVVDATIAKFGKLDIMFSNAGIG KSTISILDVDYDIKTFVDNIVGAFCAKHAARVIMPFKRSIIFTAASATVSGI VPHAYSASKGAVLGFSGKNIGVELGKGVKNCVSPHYISTPLVNLGIAERIAEKW FAAGNLKGLALLDEEVAKAVLYLASDSDSKYVSGMNLVIDGCFSTTNVALTEAYNKL F SSENTHQC"  
497..529  
misc\_feature  
/note="family signature"  
BASE COUNT 294 a 162 c 214 g 263 t  
ORIGIN  
Query Match 31.8%; Score 260.8; DB 8; Length 933;  
Best Local Similarity 57.9%; Pred. No. 8.2e-56;  
Matches 463; Conservative 0; Mismatches 337; Indels 0; Gaps 0;

QY	7	CTTCGAAGTCATTCGCAAGAGGCTAGAGAGAAAGTTGCCCTTTATAACAGGAGGACC	66
Db	20	CTTCCATACCAATAGCAAAAAGGCTAGAGGTAAGTAGCACTAATAACTGGTGAGCT	79
QY	67	AGTGGAAATTCGACAAACACAGCAAAACTCTCTCCCAACATGAGCCAAAGTTGCCATT	126
Db	80	AGCGCATAGGACAGCCAGCTAGCTTTTGTTCACATGTCGAAGGTTTACAATT	139
QY	127	GCTGATGTCGAAGATGAATTAGTCACTCAGTTGTGCGAGGCCATTGGCACTTCCAATTCC	186
Db	140	GCAGACATTCAAGACAAACCTTGCAACTCTTTTAGTACAAGAAATTCGCAACGAACATACA	199
QY	187	ACCTACATCCACTGCTGATTTACTATATGAAGAGCGGTGTTAAATAATGCCGTGGACACACA	246
Db	200	ATCTTGTTCATTCGAATGTGCTGTGTATGATCAGACGTTTCAAAATGTTAGATGCAACA	259
QY	247	GTTTAACTATGGAACATGGACATTATGTTCCAGCAATGCAGGAATTTCTGATCCCAAC	306
Db	260	ATTGCCAAATTTGGTAGCTGGACATAATGTTTCAGTAACGCTGGTATAGGAGGTAAGTCA	319
QY	307	AGGCCCGGCATCATAGACAAACGAAAGCAGACTTTTGAACGCGCTTCTCAGTGTGTAATGTA	366
Db	320	ATTTCAGCATCTTAGATGTCGATTACGACATAATTAAGACAGTGTTCGATGTAAACATT	379
QY	367	ACCGGAGTTTTCATGATGATGAGCAGGACGACGTTGTTATGATTCACAGCACCGCTGGC	426
Db	380	GTTGGCGCCTCTTTTTCGGCTAAGCATGTGCTAGAGTAATGATTCATTCAGAAAGGT	439
QY	427	AACATAATTTCCACTGCTAGTTTAAAGCTCAACTATGGTGGTGGTTCACATGCTTAT	486
Db	440	TCCATTATCTTCACAGAGTGTGCGACAGTGGTCTCTGGTATTTGTCGCGACGCGATAT	499
QY	487	TGTGGTTCAAAGCATGCTGTGTAGCCCTTACTPAGGAATCTGCGAGCTCGAGCTCGGACAA	546
Db	500	TCGTCATCAAAAGTGCAGTTTGGGATTTTCCAGAACATTTGGGTGCGAATTAGGGAAG	559
QY	547	TTTGGCATAGGTTTAATGTGTCTCTCTTCGCGCTTCTCAGCGCTTTAGGCAAGAAA	606
Db	560	TACGGAATAAAGTTAACTGTGTTTCTCCTCATTTACATTAGCACACCACTTGTATTGAAT	619
QY	607	TTTTACGGGATTAATAATGAAGAAGTAATGAGAATGTAATAACTTTGCGGGAATTTG	666
Db	620	GCCTTGGAACTAGAGAGAGATAGCGGAAATAATGTTTCCAGCAGAGGAGAAATTTG	679
QY	667	AAAGGTCCAAATTAATGTGAGGATGTTGCCAATGCAGCTCTTATCTGGCTAGTGTAT	726
Db	680	AAAGGAGCTTTACTAGATGAAGAGGAGTTCGCAAGCGAGTGTCTATACATTAGCAAGTAT	739
QY	727	GAGGCAAAATACGTGATGAGCACAATCTGTTCAATTGATGGAGGTTTACCGCTCTCAAT	786
Db	740	GATTCATAATATGATGATGGTATGAATCTGGTTATTGATGGTGGTTTTAGTACCAAAAT	799
QY	787	TCGTAAATCAAGTGTTCCTCA	806
Db	800	GTGGCTTTAACAGAGGCGCTA	819
RESULT	3		
LOCUS	AF286650	930 bp	mRNA linear PLN 02-AUG-2001
DEFINITION	Cucumis sativus CTA (CTA) mRNA, complete cds.		
ACCESSION	AF286650		
VERSION	AF286650.1	GI:15077027	
KEYWORDS	Cucumis sativus.		
SOURCE	Cucumis sativus.		
ORGANISM	Cucumis sativus		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Rosidae; eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.		
REFERENCE	1 (bases 1 to 930)		
AUTHORS	Sun, J.-Q., Li, Q.-Z., Li, X.-G. and Zhang, X.-S.		
TITLE	Gene cloning and expression of CTA in cucumber		
JOURNAL	Unpublished		

REFERENCE	2 (bases 1 to 930)		
AUTHORS	Li, Q.-Z., Li, X.-G., Sun, J.-Q. and Zhang, X.-S.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-JUL-2000) College of Life Sciences, Shandong Agricultural University, 61 Daizong Street, Taitan, Shandong 271018, P.R. China		
FEATURES	Location/Qualifiers		
source	1..930		
gene	/organism="Cucumis sativus"		
	/db_xref="taxon:3659"		
CDS	1..930		
	/gene="CTA"		
	22..837		
	/gene="CTA"		
	/note="short chain alcohol dehydrogenase; similar to TASSELSEED2"		
	/codon_start=1		
	/product="CTA"		
	/protein_id="AAK83035.1"		
	/db_xref="GI:15077028"		
	/translation="MSIQLLPAIARRLEKQVAVITGGARGIGEQTAKLFFHKHAKVVI		
	ADIQHLGQTLCKDLGSSVDFVHCDVTKEKDVETAVDTFAVSKYGLKLDLMLNAGVFE		
	ESPFDILKDDPLTFQVVNVLVGASLGRHAARVMKPAGRGSIIVTTASICSGIGI		
	GTHATSSKHVGLGLMRNAAVDLGRYIRVNCVSPNVPTENGRLKFLKVDGGEPPSF		
	YWSLKNQDILREEDVGEAVVYLGSDSCVSLNLIYDGGFTVFNQALCSFRS"		
BASE COUNT	269 a 160 c 251 g 250 t		
ORIGIN			
Query Match:	31.8%; Score 260.4; DB 8; Length 930;		
Best Local Similarity	60.0%; Pred. No. 1e-55;		
Matches	473; Conservative 0; Mismatches 306; Indels 9; Gaps 2;		
QY	1	ATGCAGCTTCGAACTCGCAAGAAGCTAGAGAGAAAGTTGCCCTTTATAACAGGA	60
Db	28	ATCCAAATGCTTCCCGCCATTGCAAGAGACTTGAAGGTAAGTAGCCGTAACTACTG	87
QY	61	GGAGCCAGTGGAAATGGAGAAACCCAGCAAAACTCTTCTCCCAACATGGAGCAAA	120
Db	88	GGGCTAGAGGAATTTGGGAAACAAACAGCAAGCTCTTCTCAAGCATGGAGCAAA	147
QY	121	GCCATTGCTGATGTCGAAGATGAATTAGGTCACTCAGTTGTCGAGGCCATTGGCA	180
Db	148	GTATTTCAGACATTCAGACCACTTTAGGTCAAACTATGTAAGGATCTCGGTCA	207
QY	181	AATTCACCATACATCCACTGTGATGTTACTTAATGAAGACGGTGTAAAAATG	240
Db	208	TCTTCT	267
QY	241	AACACAGTTTCAACCTATGGAAACTGGACATTTATGTTACGCAATGCAGGAAT	297
Db	268	ACGGCAGTCTCCAAAGTACGGAAGTATGACATCATGTTGAACAATGCAGGAG	327
QY	298	GATCCCAACAGGCCCGCATATAGACAGAAAGAGACATTTGAACCGCTTCTCAG	357
Db	328	GAATCTCCAAACTTCGACATTCGAAAGATGATCCATTAACCTTTTCAGAGAT	387
QY	358	GTAATGTAACCGGAGTTTTCCTTATGCATGAAGCAGCAGCAGCTGTATGAT	417
Db	388	GTCAACCTGGTTGGGCTCTCTCGGAACGAGACACGACGACGAGTAATGNA	447
QY	418	CGCAGTGGCAACAATAATTTCCACTGTAGTTTAAGTCAACTATGGGTGGT	477
Db	448	GGTCGAGGGAGCATCTGACGACGCGAGTATATGTCGGTATTTGGGGAAT	507
QY	478	CATGCCATTTGTTGGTTCARAAGCATGCTGTTAGCCCTTACTAGGAATCT	537
Db	508	CATGCTTATACAAGCTCGAAGCATGGGTGTTGGGATGATGAGGAATGCAG	567
QY	538	CTCGGACAAATTTGGCATTAGGCTTAATTTGTTGCTCTCTTTCGGGCTT	597
Db	568	TTGGGAGATATGGGATAGGGTAAATTTGTTTCCACCAATGTAGTCCAACT	627
QY	598	GGCAAGAAATTTTCAGGGATTAATAAATGAAGAAGATTTGAGAATGTAATA	657

```

Db 628 GGAAGAAAGTGTTCAGAGTTAAAGATGGTGGGGAATTTCCAAAGTTT-----CTACTCG 681
QY 658 GGAATTTTGAAGGTCCTCAAAATTTAATGTTGAGGATGTTGCCAATGCCAGCTCTTTATCTG 717
Db 682 AGTTTAAAAAATGAGATATTTAAAGGAAGAAGATGTGGTGAAGCTGTTGTGTAATTTG 741
QY 718 GCTAGTGTATGAGGAAAATACGTGAGTGGACACAATCTGTTCAATGATGGAGGTTTCAGC 777
Db 742 GGGAGTGCAGAGTCAAAGTGTGTGAGTGGACTCAACTTGATTGTTGATGGAGGCTTTACT 801
QY 778 GTCTGCAA 785
Db 802 GTTGTCAA 809

RESULT 4
AF286651
LOCUS Cucumis sativus TASSELSEED2-like protein (CTS2) mRNA, complete cds.
ACCESSION AF286651
VERSION AF286651.1 GI:15077029
KEYWORDS Cucumis sativus.
SOURCE Cucumis sativus.
ORGANISM Cucumis sativus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
1 (bases 1 to 930)
Sun, J.-Q., Li, Q.-Z. and Zhang, X.-S.
Functional analysis of CTS2 in cucumber
2 (bases 1 to 930)
Sun, J.-Q., Li, Q.-Z. and Zhang, X.-S.
Direct Submission
Submitted (13-JUL-2000) College of Life Sciences, Shandong
Agricultural University, 61 Daizong Street, Taian, Shandong 271018,
P. R. China
Location/Qualifiers
1. 930
/organism="Cucumis sativus"
/db_xref="taxon:3659"
1. 930
/gene="CTS2"
22..837
/gene="CTS2"
/codon_start=1
/product="TASSELSEED2-like protein"
/protein_id="AK83036.1"
/db_xref="GI:15077030"
/translation="MSIQLLPAIARRLEGKAVITGGARGIGETAKLFKHKAGVVI
ADIDHLGOTLCKDLGSSSEVHCDDVTKEKDVETAVDTAVSKYKLDIMLNAGVFE
ESNFDLKDPLITQVVVNLVGLFGRKHAARVMKPAKRGSIIVTASICSVIGI
GTHAYTSKHGVLGLMNAADVLRGVNCPNVVPTVMGRKLFKVDGGEFSPF
YWSLKNGLDILREEDVGEAVVYLGSDSKVCSGLNLIIVDGGFTVVNQALCSFRS"
BASE COUNT 270 a 158 c 249 g 253 t
ORIGIN
Query Match 31.8%; Score 260.4; DB 8; Length 930;
Best Local Similarity 60.0%; Pred. No. le-55;
Matches 473; Conservative 0; Mismatches 306; Indels 9; Gaps 2;
QY 1 ATGCAGCTCGAATCGATTCGACGAAGCTYAGAGGAAAAGTTGCCCTTATACAGCA 60
Db 28 ATCCATTTGCTCCCGCATTCGACGAAGACTTGAAGGTAAAGTAGCGTAAATCACTGGT 87
QY 61 GGAGCCAGTGAATTTGAGAAAACACAGCAAAACTCTTCCCAACATGGAGCCAAAGTT 120
Db 88 GGGGCTAGAGAAATTTGGGGAACAAACAGCGAAGCTCTTCTTCAAGCATGGAGCCAAAGTG 147
QY 121 GCCATTGCTGATGCCAAGATGAATTAGGTCTACTAGTTGTCTCGAGGCCATTGGCACTTCC 180
Db 148 GTTATTGCAGACATTCAAGACCATTTAGGTCAAACACTATGTAAGGATCTCGTCACTCA 207

```

```

QY 181 AATTCACCTACATCCACTGATGTTTACTAATGAAGACGGTGTAAAAATGCCGTGAC 240
Db 208 TCTTCTGTCTTCGTTCAITTCGACGCTAAACAAAGAGAACGTTGAACCTCGCGTTGAC 267
QY 241 AACACAGTTTCAACCTATGGAAGAACTGGACATTTATGTTTCAGCAATGCAGGAATTT--CT 297
Db 268 ACGGCAGTCTCCAAGTACGGAAGATTAGACATCATGTTGAACATTCAGAGTGGTGAAC 327
QY 298 GATCCACAGAGCCCGGATCATAGACAAACGAAACAGACACTTTGAACGGTTCACAGT 357
Db 328 GAATCTCCAAACTTCGACTTTCTGAAAGATGTTCCATTAACTTTTCAGAGTGGTGAAC 387
QY 358 GTAATGTAAACCGGAGTTTTCCTATGCATGAAGCAGCAGCAGCTGTATGATTCACGCA 417
Db 388 GTCAACCTGGTTGGGGCCCTTTCTCGGAACGAAACACGACGACGAGTAATGAACCCAGCG 447
QY 418 CGCAGTGGCAACATATAATTTCCACTGCTAGTTTAAAGCTCAACTATGGGTGGTCTTTCA 477
Db 448 GGTGAGGAGGAGCATGTCGACGACAGCAGATATGCTCGGTGATTGGGGGAATTTGGGACG 507
QY 478 CATGCTATTGTTGTTTCAAAGCATGCTGTGTAGCCCTTACTAGGAATCTGGCAGTCGAG 537
Db 508 CATGCTTATACAAGCTCGAAGCATGGGGTGTGGGATTGATGAGGAATGCAGCTGGAT 567
QY 538 CTCGGACATTTTGGCATTAGGTTAATGTTTGTCTCTCTTCGGGCTTCCTACGGCTTTA 597
Db 568 TTGGGAAGATATGGGATTAGGTTAATTTGTTTTCACCAATGTAGTGCACCACTGAATG 627
QY 598 GGCAGAAATTTTTCAGGATTAATAATCAAGAAGAAATTTGAGAATGTAATAAATCTTTCG 657
Db 628 GGAAGGAAGTTGTTTCAAGGTTAAAGATGGTGGGAAATTTCCAAAGTTT-----CTACTGG 681
QY 658 GGAATTTTGAAGGTCCTCAAAATTTAATGTTGAGGATGTTGCCAATGCAGCTCTTTATCTG 717
Db 682 AGTTTAAAAATGGAGATATTTTAAAGGGAAGAAGATGTGGGTGAAGCTGTGTGCTATTG 741
QY 718 GCTAGTGTAGGAGCAAAATACGTGAGTGCACACAATCTGTTTCATTTGATGAGGTTTCAGC 777
Db 742 GGGAGTGCAGAGTCCAAGTGTGTGAGTGGACTCAACTTGATTGTTGATGAGGCTTTACT 801
QY 778 GTCTGCAA 785
Db 802 GTTGTCAA 809

RESULT 5
AF072447
LOCUS Ipomoea trifida short-chain alcohol dehydrogenase (SSP) mRNA, SSP-1
DEFINITION Ipomoea trifida short-chain alcohol dehydrogenase (SSP) mRNA, SSP-1
ACCESSION AF072447
VERSION AF072447.1 GI:3598856
KEYWORDS Ipomoea trifida.
SOURCE Ipomoea trifida.
ORGANISM Ipomoea trifida
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.
1 (bases 1 to 1089)
Koyama, Y., Kadota, N., Sassa, H., Kakeda, K., Tsuchiya, T.,
Ishimizu, T., Kondo, K. and Norioka, S.
Self-incompatibility protein (SSP) from sporophytic
short-chain alcohol dehydrogenase family
Unpublished
2 (bases 1 to 1089)
Koyama, Y. and Kadota, N.
Direct Submission
Submitted (18-JUN-1998) Faculty of Bioresources, Mie University,
Kamihama-cho, Tsu, Mie 514-8507, Japan
Location/Qualifiers
1. 1089
source

```



```
/organism="Ipomoea trifida"
/db_xref="taxon:35884"
/note="wild diploid species related to sweet potato"
1. .1089
/gene="SSP"
/allele="SSP-1"
44. .898
/gene="SSP"
/note="self-incompatibility-locus specific stigma protein"
/codon_start=1
/product="short-chain alcohol dehydrogenase"
/protein_id="AAC35340.1"
/db_xref="GI:3598857"
/translation="MAASLQPGSKRLGKVAITGAANGIGATTARLFAHQGCKVI
IADIDKNGHVAEIGVEYALFICHDVIESDVQNAVDTTVSRVKGKLDIMFSNAGVI
GSRDTSILEASPENILVFETNVGAFFCAKHAARVMIPARKSGVIFASASAEEVGI
TSDTTSKCAVVGKSLICVMKGYKIKANCVSPVILTKLGMSIMPTQDPKLAEEI
VAEASNFGKLTITEDVEAALYLAGDESKFVSGNLNLLDGGFTTTNIAFQVAVEKVL
GGGEETTA"
BASE COUNT 303 a 249 c 282 g 255 t
ORIGIN
Query Match 30.8%; Score 252.6; DB 8; Length 1089;
Best Local Similarity 57.4%; Pred. No. 1e-53; Indels 0; Gaps 0;
Matches 456; Conservative 0; Mismatches 339;
QY 7 CTTCTGAACCTGCAATTCGCAAGAGGCTAGAGAAAGTTGCCCTTATAACAGGAGGAGCC 66
Db 59 CTTCAACTCCCGGTCCTCAAGACACTAGAGGAAGGTTGCTATATAACAGCGCGAGCC 118
QY 67 AGTGGAAATGGAGAAACACACAGCAAACTCTTCTCCCAACTGAGGCCAAAGTTGCCATT 126
Db 119 AACGGCATAGGAGCAACACCGCGAGGCTCTTTGCTCAACATGGCTGCAAGGTTATCATT 178
QY 127 GCTGATGCTCAAGATGAATAGTCTACTCAGTTGTGAGGCCATTGGCAGCTTCCATTCC 186
Db 179 GCTGACATGACCAAGAAGATGGCCACTCGTGGCCGAGAGATCGGCCGGAATACGCA 238
QY 187 ACCTACATCCACTGTGATGTTACTAATGAAGACGGTGTAAATAATGCCGTGGACACACA 246
Db 239 CTATTCATCCACTGCGAGCTCAGAAATCGAATCAGAGCTCAAAATCGGTGGACACACC 298
QY 247 GTTTCACCTATGGAACCTGGAATATTATGTCAGCAATGCGAGGAATTTCTGATCCCAAC 306
Db 299 GTTTCAGGTACGGAAGTTAGACATCATGTTTCAGCAATGCGAGCGCTCGCAGGAGCGCG 358
QY 307 AGCCCGCGCATCATACACACGAAAGACAGACTTTGAACGCGTTCTCACTGTAATGTA 366
Db 359 GACACAGCATCTGGAGCGCCAGCCAGAAAACATCAACCTAGTTTTCGAGAGCAATGTG 418
QY 367 ACCGAGATTTTCTATGTCATGAAGCAGCAGCAGCTGTTATGATTCAGCAGCAGCTGGC 426
Db 419 TTCGAGCATCTTCTGCGCCAAACACGCGGCAAGGTTGATGATCCGCGAGGAGGCG 478
QY 427 ACATATATTTCCACTGCTAGTTTAAGCTCAACTATGGGTGGTGGTTCCTTCACATGCCTAT 486
Db 479 AGCGTAATATTTCTCCGAGCGCGCGCTCGGAGGTCTCGGATCATCCTCCGACAGCTAC 538
QY 487 TGTGGTTCAAAGCATGCTGTGTAGCCCTTACTAGGAATCTGCAGCTGAGCTCGGACAA 546
Db 539 ACGGCATCAAGTGCCTGCTGCGGACTATGCAAGAGCGCTTGCCTTGAATGGGAAA 598
QY 547 TTTGGCATTAGGTTAATGTTGTCTCTCTCGGCTTCTTACGCTTTAGGCAAGAAA 606
Db 599 TATGGATCAAGCCAACTTCGCTTTCGCTTATGTAATTTCTGACTAAGTTGGGAATGAGT 658
QY 607 TTTTCAGGATTAAAAATGAAGAAGAAATTTGAGAATGTATAAATCTTTCGCGGAAATTTG 666
Db 659 ATAATGCCCAACCCAGGATAGGAAATGGCGGAGGAAATTTGTGCGAAGCTTCGAATTC 718
QY 667 AAGGTCCAAAATTTAATGTTGAGGATGTTGCCAATGAGCTCTTTATCTGGCTAGTGAT 726
Db 719 AAAGGGAAAACCTTTGACGACAGAGGATGAGCAGAGCGGCTTTGTACCTCGCGCGGAC 778
```

```
QY 727 GAGGCAAAATACGTGAGTGGACACAAATCTTTCATTGATGAGGTTTCAGCGTCTCAAT 786
Db 779 GAGTCTAAGTTCGTAAGCGGCTCAACCTTTGATTGATGGGGGATTTACCACCACCAAT 838
QY 787 TCTGTAATCAAGTG 801
Db 839 ATAGCCTTTCAAGTG 853
RESULT 6
NTAJ3178 Nicotiana tabacum SCANT gene. 1152 bp DNA linear PLN 30-JAN-1998
LOCUS AJ223178
DEFINITION AJ223178
ACCESSION AJ223178
VERSION AJ223178.1 GI:2791347
KEYWORDS SCANT gene; short chain alcohol dehydrogenase.
SOURCE Nicotiana tabacum.
ORGANISM Nicotiana tabacum
REFERENCE 1 (bases 1 to 1152)
AUTHORS Herbers,K., Monke,G., Badur,R. and Sonnewald,U.
TITLE A simplified procedure for the subtractive cDNA cloning of
photoassimilate-responder genes: isolation of cDNAs encoding a new
class of pathogenesis-related proteins
JOURNAL Plant Mol. Biol. 29 (5), 1027-1038 (1995)
MEDLINE 96145513
PUBMED 8553446
REFERENCE 2 (bases 1 to 1152)
AUTHORS Moenke,G.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1997) Moenke G., Molecular Genetics, Institut
fuer Pflanzenzuenetik und Kulturpflanzenforschung, Gatersleben,
D-06466, GERMANY
FEATURES
Source
1. .1152
/organism="Nicotiana tabacum"
/variety="SNN"
/db_xref="taxon:4097"
/tissue_type="leaf"
Join(5. .43,134. .949)
/gene="SCANT"
Join(5. .43,134. .949)
/gene="SCANT"
Join(5. .43,134. .949)
/gene="SCANT"
/codon_start=1
/product="short chain alcohol dehydrogenase"
/protein_id="CA11154.1"
/db_xref="GI:2791348"
/db_xref="SPTREMBL:O50038"
/translation="MASYSLSPSPKARLEKGVKALITGGASGIGATARLFVQHGAKVT
IADIDNGLTSLVGEIHEHTFVHCNVAVSDVQNAVDTTAKFKGLDIMFSNAGIG
KSISSILDVDYDIITKTFDVIIVGAFCAKHAARVMIPFKKGSIIFFASAATVVSIGI
VPHASGASKAVLGSFKNIGVKGKIVKNCVSPHYISTPLVLNAGTAEREIAEKW
FAAGCNLKGALLDEEVAKAVLYLASDDSKYVSGNMLYIDGFGSTTNVATLEYNKLK
SSEHTHQ"
<5. .43
/gene="SCANT"
/number=1
44. .133
/gene="SCANT"
/number=1
134. .949
/gene="SCANT"
/number=2
587. .619
/gene="SCANT"
/note="family signature"
BASE COUNT 356 a 196 c 244 g 356 t
ORIGIN
```

```
Query Match      30.8%; Score 252.6; DB 8; Length 1152;
Best Local Similarity 57.8%; Pred. No. 1e-53;
Matches 450; Conservative 0; Mismatches 329; Indels 0; Gaps 0;

QY 28 AGGCTAGAAGAAAGTTGCCCTTATAACAGGAGGAGGACGAGTGAATTTGGAGAAACACA 87
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 131 AGGCTAGAAGTAAAGTACACTTAATTAACCTGTGGAGTACGCGCATAGGAGCAGCCACA 190
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 88 GCAAACTCTTCTCCCAATGAGCAAGTTGCCATTGTGTATGTCCTCAAGATGAATTA 147
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 191 GCTAGGCTTTTGTTCACATGTGCAAGGTTTACAATTCAGAGACATTCAGCAACCTT 250
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 148 GTCACCTCAGTTGTTCAGGCGCATTTGCCACTTCCAAATTCACCTACATCCACTGTGATGTT 207
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 251 GGAACCTCTTTAGTACAGAAATTTGCCACGACATACAACTTTTGTCCATTCGAATGTT 310
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 208 ACTAATGAACAGCGTGTAAATAATGCCGTGGACACACAGTTCCTCACTATGGAACACTG 267
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 311 GCTGTTGAATCAGCGTTCAAATGTTGTAGATGCAACAAATTTGCCAAATTTGGTAAAGCTG 370
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 268 GACATTTATGTTACGCAATGAGCAATTTCTGATCCCAACAGCCCGCATCATAGACAAC 327
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 371 GACATAATGTTTCAGTAACTGCTGTATAGGAGTAACTCAATTTCCAGCATCTTAGATGTC 430
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 328 GAAAAGCAGACATTTTGAACGCGCTCTCAGTGTAAATGTAACCGAGTTCCTCATGATG 387
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 431 GATTACGACATAATTAAGACAGTGTTCGATGTAACATTTGTTGGCGCTTCTTTTGGCGCT 490
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 388 AAGCAGCAGCAGCGTGTATGATTCAGCAGCAGTGGCAACATAATTTCCACTGCTAGT 447
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 491 AAGCATGCTCTAGAGTAATGATTCATTCAGAAAGGTTCCATTTATCTTCACAGCAAGT 550
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 448 TTAAGCTCAACTAGTGGTGTGTTCTTTCACATGCTATGTTGTTTCAAGAGCATGCTG 507
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 551 GCTCGACAGTGTCTCTGTTATGTCGCGACGATATTCGTCATCAAAAGTGCAGTT 610
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 508 TTACCCCTTACTAGGAATCTGGCAGTCGAGTCGGAACAATTTGGCAATTTAGGTTAATGT 567
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 611 TTGGATTTTCCCAAGAACATTTGGGGTGCAGTAATTAGGGAAGTACGGAATAAAGTTAATCTG 670
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 568 TTGTCCTTTTCGGGCTTCTCAGCGCTTTAGGCAAGAAATTTTCAGGATTAATAATGAA 627
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 671 GTTCTCTCTATTACATTTAGCACACCACCTTTGTTATTTGCGCTTGAATAGCTGAGAGA 730
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 628 GAAGAATTTTGAGATGTAATAAATTTGCGGGAATTTGAAAGGTTCCAAAATTTAATGTT 687
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 731 GAGATAGCGAAAAATGTTTTCAGCAGGAGGAGAAATTTGAAAGCAGCTTTTACTAGATGAA 790
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 688 GAGATGTTGCCAATGACGCTCTTTATCTGCTAGTATGAGGCAAAATACGTGAGTGA 747
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 791 GAGAGTTCGCAAGCAGTGTCTATACCTAGCAAGTATGATTTCTAAATATGAGTGTG 850
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 748 CACAATCTGTTCAATGATGAGGTTTCAGCGTCTGCAATTTCTGTAATCAAAAGTTTCCA 806
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 851 ATGAATCTGTTATGATGGTGTGTTTATGACCACAAAATGCGCTTTTAAACAGAGGCCTA 909
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
D88121
LOCUS      Vigna unguiculata 1054 bp mRNA linear PLN 07-FEB-1999
DEFINITION Vigna unguiculata mRNA for CPRD12 protein, complete cds.
ACCESSION  D88121
VERSION    D88121.1 gi:1853967
KEYWORDS   CPRD12 protein,
           Vigna unguiculata
SOURCE     Vigna unguiculata one-month-old cDNA to mRNA.
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
           Vigna.
REFERENCE  1 (sites)
AUTHORS    Iuchi,S., Yamaguchi-Shinozaki,K., Urao,T. and Shinozaki,K.
```

```
Characterization of two cDNAs for novel drought-tolerant genes in
the highly drought-tolerant cowpea
J. Plant Res. 109, 415-424 (1996)
2 (sites)
Iuchi,S., Yamaguchi-Shinozaki,K., Urao,T. and Shinozaki,K.
Characterization of two cDNAs for novel drought-tolerant genes in
the highly drought-tolerant cowpea
Unpublished
3 (bases 1 to 1064)
Iuchi,S.
Direct Submission
Submitted (27-SEP-1996) Satoshi Iuchi, The Institute of Physical
and Chemical Research(RIKEN), Laboratory of Plant Molecular
Biology, 3-1-1 Koyadai, Tsukuba, Ibaraki 305, Japan
(E-mail:iuchi@rtc.riken.go.jp, tel.:0298-36-4359, fax:0298-36-9060)
Location/Qualifiers
1..1064
/organism="Vigna unguiculata"
/db_xref="taxon:3917"
/dev_stage="one-month-old"
76..879
/codon_start=1
/product="CPRD12 protein"
/protein_id="BAAL3541.1"
/db_xref="gi:1853968"
/translation="MANGSVLSVPVKRLGKVAITGGASGIGETARLFQSHGAHV
VADIQDDVGLSLCNELKSALYVHCDVTKEEDTEKVDYAVSKFGLDMFNAGTDE
FKKTSIDNKTDFERVISVNLVGPFLGTRKHAARVMPARRGCIINTASVAGCIGGAT
HAYTSSKHALVGLFKNTAFLVGLVGFIRVNCVSPFAIVTPLLNKYINLDEGVKRTYNN
LKWVYVPNDVAEAAALYLASDESKFVSSHNLVIDGGLNSNVGPFPMFM"
polya_site 1064
/note="18 A nucleotides"
BASE COUNT 311 a 198 c 254 g 301 t
ORIGIN
Query Match      30.6%; Score 250.4; DB 8; Length 1064;
Best Local Similarity 59.3%; Pred. No. 3.7e-53;
Matches 472; Conservative 0; Mismatches 306; Indels 18; Gaps 2;
QY 8 TTCGAATGCATTCGCAAGAGGGCTAGAGGAAAGTTGCCCTTATAACAGGAGGAGCCA 67
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 92 TTCTCTCAGCTGTTGTTAAAGGCTTGAGGGAAAGTGGCGGATTCACCGCGGTGCCA 151
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 68 GTGGAATTGGAGAAACACAGCAAAACCTCTTCTCCCAACATGGAGGCAAAAGTTGCCATTG 127
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152 GCGGCATAGTGTGAGGCCACTGCAAGACTCTTCTCACAGCATGGAGCTCAGCTGTGTGG 211
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 128 CTGATGTCGAAGATGAATTAGTCTCAGTTGTCGAGGCCATTTGGCACTTCCAATTCCA 187
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 212 CTGATATTCAACAGCATGTGGGACTTCTCTCTCAATGAGTTGAAATC-----AGCCA 265
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 188 CCTACATCCACTGTGATGTTACTAATGAAGACGGTGTAAATATGCCGTGGGACACACAG 247
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 266 TATATGTTTCATTCGCGATGTGACAAAGGAAGAGACATCGAAAGAGTGGTGGACACAGCA 325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 248 TTTCAACCTATGGAACACTGGACATTTATGTTTCAGCAATTCAGGAAATTTCTGTATCCCAACA 307
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 326 TTTCCAAGTTCGGAAGCTAGACATCATGTTTAATAACGCGGCTACAGGTGATGAGTTCA 385
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 308 GGGCCCGCATATAGACAAACAAAAAGCAGACTTTTGAACGCTTCTCAGGTAAATGTAA 367
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 386 AAAAAAGCATCTCGACAAACAAAAATCTGATTTTGAAGAGTGAATAGTGTGAACCTGG 445
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 368 CCGGAGTTTTCCTATGTCATGAGCAGCAGCAGCTGTATGATTCAGCAGCAGCGAGCA 427
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 446 TTGCTCCATTTCTGGGAACAAGACGCTGCAAGTCATGATCTCTCTAGTAGAGAGGT 505
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 428 ACATAATTTCCACTGCTAGTTTAACTCAACTATGGGTGGTGTCTTTCACATGCCCTATT 487
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 506 GCATAATTAACACAGCGAGTGTTCAGATGTCATAGGTGGAGGAGCTACACATGCTAC 565
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 488 GTGGTTCAAGCATCTGTTAGCCCTTACTAGGAATCTGCGAGCTCGAGCTCGGCAAT 547
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 566 CcAGTTCcAAAGcATGccCTAGTGGGAcTcACcAAAAAcAcTcGCGGTGAGcCTTGGAcAAAT 625  
 QY 548 TTGGcATTAGGGTTAAATGTTGTTCTCTCTCGGGTTCTcACGGCTTTAGGcAAAGAAAT 607  
 Db 626 TCGcATTCTGTGAAcTGTGTCTCCCTTTTGGcATTGTGAcACcCTGTGTGAAcAAAT 685  
 QY 608 TTTcAGGGATTAAATAAAGAAAGAAATTTGAGAAATGTAATAAAcTTTGGcGGGAAATTTGA 667  
 Db 686 ATTTCATcCTTTGATGAAGAAGAGAGTTGGcAAGAcCTTATATGAAC-----CTAA 733  
 QY 668 AAGcTCCAAATTTAAATGTTGAGcATGTTCCcAAATGCAGcCTCTTATCTCGcTAGTcATG 727  
 Db 734 AAGcTGGTATCTGTcGCCCTAAAGATGTGGcAGAAGcCTGTCTTACTTTAGcAAGTcATG 793  
 QY 728 AGcCAAAATAcGTGAGTGGcACAAATcTGTTCATTTGATGAGGcGTTcAGcGCTGTcCAAT 787  
 Db 794 AGcTAAATGTTGTTAGTCTCATAAATcTGTCTAGAcGGAGcGTTTcATCAATTCAAAcG 853  
 QY 788 CTGTAATCAAGTGT 803  
 Db 854 TAGGATTTCCTATGTT 869  
  
 RESULT 8  
 AF072449  
 LOCUS  
 DEFINITION  
 Ipomoea trifida short-chain alcohol dehydrogenase (SSP) mRNA, PLN 13-SEP-1998  
 ACCESSION  
 AF072449  
 VERSION  
 AF072449.1 GI:3598860  
 KEYWORDS  
 Ipomoea trifida.  
 ORGANISM  
 Ipomoea trifida  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; eusterids I; Solanales; Convolvulaceae; Ipomoea.  
 1 (bases 1 to 1179)  
 Kowama,Y., Kadota,N., Sassa,H., Kakeda,K., Tsuchiya,T.,  
 Ishimizu,T., Kondo,K. and Norioka,S.  
 S-locus specific stigma protein (SSP) from sporophytic  
 self-incompatibility plant, Ipomoea trifida, is a member of  
 short-chain alcohol dehydrogenase family  
 Unpublished  
 2 (bases 1 to 1179)  
 Kowama,Y. and Kadota,N.  
 Direct Submission  
 Submitted (18-JUN-1998) Faculty of Bioresources, Mie University,  
 Kamihama-cho, Tsu, Mie 514-8507, Japan  
 FEATURES  
 Location/Qualifiers  
 1..1179  
 /organism="Ipomoea trifida"  
 /db\_xref="taxon:35984"  
 /note="wild diploid species related to sweet potato"  
 1..1179  
 /gene="SSP"  
 /allele="SSP-10"  
 47..880  
 /gene="SSP"  
 /note="self-incompatibility-locus specific stigma protein"  
 /codon\_start=1  
 /product="short-chain alcohol dehydrogenase"  
 /protein\_id="AAC35342.1"  
 /db\_xref="GI:3598861"  
 /translation="NAASLQLPVFKRLGKVAIITGAANGIGATTARLFAOHGCKVI  
 IADIDKNGLSVAEEIGPEYALFIHCDVRIESDVONAVDTTVSRYGKLDIMFSNAGVA  
 GSRDTSILEASPENINLPETNVFGAFCAKHAARWIPARKGSVIFSASAASEVFGI  
 TSDTITASKCAVVGKSLCVMKGIGKANCSPYVILTKIGMSIMPTQDRKLAEEI  
 G"  
 348 a 251 c 290 g 290 t  
 BASE COUNT  
 ORIGIN  
 30.1%; Score 246.2; DB 8; Length 1179;

Best Local Similarity 56.9%; Pred. No. 4.3e-52;  
 Matches 452; Conservative 0; Mismatches 343; Indels 0; Gaps 0;  
 QY 7 CTTTGAACGTCTCCCAAGAGCTAGAGAGAAAGTTGCCCTTATACAGAGAGAGCC 66  
 Db 62 CTTTCAACTCCCGCTGTTCAAAAGACTAGAAAGAAAGTGTCTATCATACAGAGAGAGCC 121  
 QY 67 AGTGAATTTGGAGAAACCACAGCAAAACCTCTCTCCCAACATCGAGCCAAAGTTGCCATT 126  
 Db 122 AACGGCATAGGAGCAACACCGCGAGGCTCTTTGCTCAACATGGCTGCAAGGTCATCAT 181  
 QY 127 GCTGATGTCCAAAGTAAATAGTCTACTAGTGTGCGAGCCATTTGGCACTTCCAATTC 186  
 Db 182 GCTGACATTTGACGACAAGATGGCTCTCGTGGCGGAAGAGATCGGCCGGAATACGA 241  
 QY 187 ACCTACATCCACGTGATGTTACTAATGAAGACGGTGTAAATAATGCCGTGGACAACACA 246  
 Db 242 CTATTATCCACTCGCAGCTCAGATCGAATCGAATCAGAGCTCAAAATGGGTGGACACACC 301  
 QY 247 GTTTCAACCTATGGAACCTGGACATTTATGTTCAAGCAATCGAGGAATTTCTGTATCCCAAC 306  
 Db 302 GTTTCCAGGTACGGAAGCTAGACATCATGTTTCAGCAATGCAGGCGTCGCAGGGAGCCGC 361  
 QY 307 AGGCCCGGATCATAGACAAGAAAAAGACACTTTGAAGCGCTTCTCAGTGTAAATGTA 366  
 Db 362 GACACAGCATCTTTGGAGGCCAGCCAGAGAACATCAACCTAGTCTTCGAGAGCAATGTG 421  
 QY 367 ACCGGAGTTTTCCTATGTCATGAAGCAGCAGCAGCTGTTATGATCCAGCAGCAGTGGC 426  
 Db 422 TTCGGAGCTATTCTTCGCGCAACACGCGCGAGGTGATGATCCCGCGAGGAAGGGC 481  
 QY 427 AACATAATTTCCACTGCTAGTCTTAAAGCTCAACTATGGTGGTGGTCTTCACATGGCTAT 486  
 Db 482 AGCGTAATATCTCGGGAGCGCGCGCTCGGAGTGTTCGGGATCACTCCGACACGTAC 541  
 QY 487 TGTGGTTCAAAGCATGCTGTGTAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGACAA 546  
 Db 542 ACGGCATCCAAGTCGCGCGCTCGTGGGACTATGAAGAGCCTGTCCGTGAAATGGGAAA 601  
 QY 547 TTTGGCATAGGGTTAAATGTTGTTCTCTCCCTTCGGGCTTCTCAGGCTTTAGGCAAGAA 606  
 Db 602 TATGGATCAAGCCCACTCGCTTCGCTTATGTAATCTGACTAAGTTGGGAATGAGT 661  
 QY 607 TTTTCAGGGATTAAAAATGAAGAAGAAATTTGAGAATGTAATAAACTTTGCGGAAATTTG 666  
 Db 662 ATAATGCCAGCCAGGATAGGAATTTGGCGGAGGAAATTTGCGGGAAGCTTCGAATTC 721  
 QY 667 AAAGTCCAAAATTTAATGTTGAGGATGTTGCCAATGCAGCTCTTTATCTGGCTAGTAT 726  
 Db 722 AAAGGGAAAACTTTGACGACAGAGGATGTAGCAGAGCGGCTTTGTACCTCGCGCGGAC 781  
 QY 727 GAGGCAAAATACGTGAGTGGACACAATCTGTTCAATTGATGGAGGTTTCAGCGTCTCAAT 786  
 Db 782 GAGTCTAAGTTCGTGAGCGGGCTCAACCTTCTGATTGATGGGGGATTTACCACCACCAAT 841  
 QY 787 TCTGTAATCAAAGTG 801  
 Db 842 ACAGCCTTTCAAGTG 856  
  
 RESULT 9  
 AB070221  
 LOCUS  
 DEFINITION  
 Ipomoea trifida SSP10 gene for S-Locus linked stigma protein, complete cds.  
 ACCESSION  
 AB070221  
 VERSION  
 AB070221.1 GI:19911178  
 KEYWORDS  
 Ipomoea trifida (strain:H77-2) DNA, clone\_lib:Lambda FIXII Ipomoea trifida genomic Library.  
 ORGANISM  
 Ipomoea trifida  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

```

REFERENCE
AUTHORS Tsuchiya,T., Suwabe,K., Watabe,T., Kagaya,Y. and Kowiyama,Y.
TITLE Characterization of promoter region of the SSP gene from Ipomoea
JOURNAL trífida regulates stigma-specific expression
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 3021)
TITLE Tsuchiya,T.
JOURNAL Direct Submission
SUBMITTED (15-AUG-2001) Tohru Tsuchiya, Mie University, Lab. of
Plant Molecular Genetics and Breeding; Kamihama, 1515, Tsu, Mie
514-8507, Japan (E-mail:tsuchiya@bio.mie-u.ac.jp,
Tel:81-59-231-9515, Fax:81-59-231-9515)
FEATURES
source Location/Qualifiers
1..3021
/organism="Ipomoea trifida"
/strain="H77-2"
/db_xref="taxon:35884"
/clone_lib="Lambda FIXII Ipomoea trifida genomic Library"
/notes="vector: pBluescript SK-"
1543..1550
Join(1649..1686,1787..2582)
/gene="SSP10"
Join(1649..1686,1787..2582)
/gene="SSP10"
/codon_start=1
/product="S-Locus linked stigma protein"
/protein_id="BAB86916.1"
/db_xref="GI:19911179"
/translation="MAASSLQLPVFKLEKVAITGAANGIGATTARLFAOHGCKVI
IADIDKNGLSVAEIGPEYALFHCDVRIESHVQHAVDTIVSRYGRLDITFNSAGVA
GSRDTSLEASPENINLVFETNVFGAFCAKHAARVILPARKGSVIFSAASAEEVFI
TSDYATASKAVGLKSLCVEMGKYGKANGCVSPVILTKLGMISIMPTQDRKLAEEI
VAEASNFKGLTATEDVAEAAALYLAGDESKFVSLNLLIDGGFTTTNTAFQAVKEIL
G"
polyA_signal 2631..2636
polyA_signal 2775..2780
BASE COUNT 1005 a 549 g 574 t 892 c 1 others
ORIGIN

Query Match
Best Local Similarity 29.84; Score 244.4; DB 8; Length 3021;
Matches 443; Conservative 0; Pred. No. 1.1e-51;
Mismatches 331; Indels 0; Gaps 0;

QY 28 AGCCTAGAGGAAAGTTGCCCTTATAACAGGAGGAGCCAGTGAATGGAGAACACCA 87
DB 1785 AGACTAGAGGAAAGTTGCTATCATACAGCGCGAGCCAGCGCATAGAGCAACCC 1844
QY 88 GCMAAACTCTTCCCAACATGAGGCCAAAGTTGCCATTGCTGATGCCAGATGAATTA 147
DB 1845 CGAGGCTCTTTGCTCAACATGCTGCAAGGTCAATCTGCTGATGCCAGATGAATTA 147
QY 148 GGTCACTCAGTTGCGAGGCCATTTGCCATCTTCCATTCACCTACCTCACTGTGATGT 207
DB 1905 GGCTCTCCGTCGGGAGAGATCGCGCGGAATACGCACTATTTCATCCACTCGGAGTC 1964
QY 208 ACTANTCAAGACGGTGTAAATAACCGTGGACACACAGTTCACCTATGGAACATG 267
DB 1965 AGAATCGATCAGATCGCAATATGCGGTGGACACACACCTTTCCAGGTACGGAAGCTA 2024
QY 268 GACATATCTTCAGCATGCGAGGAATTTCTGATCCCAACAGCGCGCGCATATAGCAAC 327
DB 2025 GACATATCTTCAGCATGCGAGGAATTTCTGATCCCAACAGCGCGCGCATATAGCAAC 327
QY 328 GAAAGACGACGCTTGAACGGTCTTCAGTCTAAATGTAACCGGAGTTTCTCTATGATG 387
DB 2085 AGCCGAGACACATCACTGATGTTTCGAGACGAATGTGTTCCGAGCATCTCTCTGCGCC 2144
QY 388 AAGCAGCAGCAGCTGTATGATTCAGCAGCAGCTGGCAACATATTTCCACCTGCTAGT 447
DB 2145 AAACACGGCGGCGAGGTGATGATCCCGGGGAGGAAGGGCAGCGTAAATATCTCGCGGAGC 2204
QY 448 TTAAGCTCAACTATGCGTGGTGTCTTCCATGCTGCTATTTGTTCAAGCATGCTGTG 507

```

```

Db 2205 GCGGCGTCGGAGTGTTCGGGATCACCTCCGACACGTACACGCATCCAGTGCCTGC 2264
QY 508 TTAGCCCTTACTAGGAATCTGGCAGTCGACCTCGGACAAATTTGGCATTTAGGTTAATGT 567
Db 2265 GTGGGACTATGCAAGAGCCTGTGGTGAATGGGAAATATGGGATCAAGCAACTGC 2324
QY 568 TTGCTCTCTTCGGGCTTCCTAGCGCTTTAGGCAAGAAATTTTCAGGATTAATAAGTAA 627
Db 2325 GTTTCGGCTTATGTAATCTGACTAAGTTGGGATGAGTATTAATCCCGACCCAGATAG 2384
QY 628 GAAGAATTTGAGATGTAATAAACTTTGCGGAAATTTGCAAGTTCCTCAAAATTTAATGTT 687
Db 2385 AAATTCGGCGGAGAAATTTGTCGGAAGCTTCGAATTTCAAGGGAACACTTTGACGACA 2444
QY 688 GAGATCTTCCCAATCAGCTCTTTATCTGCTAGTATGAGGCAAAATACGTGAGTGA 747
Db 2445 GAGATCTAGCAGAGCGCTTTGACTCTCGCGCGAGCTCTAAGTTTCGTGAGCGG 2504
QY 748 CACAATCTCTTCATTCATGAGGCTTCAGCGTCTGCAATCTCTTAATCAAGTG 801
Db 2505 CTCAACCTTCTGATTCGAGGGAATTTACCAACCAATACAGCTTTCAAGTG 2558

RESULT 10
AF072448 1120 bp mRNA linear PLN 13-SEP-1998
LOCUS Ipomoea trifida short-chain alcohol dehydrogenase (SSP) mRNA, SSP-9
DEFINITION Af072448
ACCESSION Af072448
VERSION Af072448.1 GI:3598858
KEYWORDS
SOURCE Ipomoea trifida.
ORGANISM Ipomoea trifida
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Eudicotyledons; Core eudicots;
Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.
1 (bases 1 to 1120)
Kowiyama,T., Kadota,N., Sassa,H., Kakeda,K., Tsuchiya,T.,
Ishimizu,T., Kondo,K. and Norioka,S.
Self-incompatible stigma protein (SSP) from sporophytic
short-chain alcohol dehydrogenase family
Unpublished
2 (bases 1 to 1120)
Kowiyama,T. and Kadota,N.
Direct Submission
Submitted (18-JUN-1998) Faculty of Bioresources, Mie University,
Kamihama-cho, Tsu, Mie 514-8507, Japan
Location/Qualifiers
1..1120
/organism="Ipomoea trifida"
/db_xref="taxon:35884"
/notes="wild diploid species related to sweet potato"
1..1120
/gene="SSP"
581..891
/allele="SSP-9"
/gene="SSP"
/notes="self-incompatibility-locus specific stigma protein"
/codon_start=1
/product="short-chain alcohol dehydrogenase"
/protein_id="AAC35341.1"
/db_xref="GI:3598859"
/translation="MAASSLQLPVFKLEKVAITGAANGIGATTARLFAOHGCKVI
IADIKNGHVAEEIGPEYALFHCDVRIESHVQHAVDTIVSRYGRLDITFNSAGVA
GSRDTSLEASPENINLVFETNVFGAFCAKHAARVILPARKGSVIFSAASAEEVFI
TSDYATASKAVGLKSLCVEMGKYGKANGCVSPVILTKLGMISIMPTQDRKLAEEI
VAEASNFKGLTATEDVAEAAALYLAGDESKFVSLNLLIDGGFTTTNTAFQAVDKIL
G"
BASE COUNT 315 a 255 c 281 g 269 t
ORIGIN

```



```

Qy 487 TGTGGTTCAAAGCATCTGTTAGCCCTTACTAGGAATCGCAGTCGAGCTCGGACAA 546
Db 554 ACGGCATCAAGTCGCGCTGCGGACTATGCAAGAGCCCTGCGCTTGAATGGGGAAA 613
Qy 547 TTTGGCATTAGGGTTAATGTTTGTCTCTTTCGGCTTCCCTACGGCTTTAGGCAAGAAA 606
Db 614 TATGGGATCAAGCCAACTGCGTTTCGCCCTATGTAATCTGACCAAGCTGGGAATGAGT 673
Qy 607 TTTTCAGGATTAATAATGAAGAGAAATTTGAGAATGTAATAAATTTGCGGGAAATTTG 666
Db 674 ATAATGCCGACCCAGGATAGAAATTCGCGGAGGAAATTTGTCGCGAAGCTTCGAATTTT 733
Qy 667 AAAGTCCAAAATTAATGTTGAGGATGTTGCCAATGCAAGCTCTTTATCTGCTAGTGAT 726
Db 734 AAAGGAAATTTTACGACGGAGGATGACGAGGGCGCTTTGTACTCGCGCGGAC 793
Qy 727 GAGGCAAAATACGTGAGTGACACAACTGTTTCATGATGAGGAGGTTTCAGCGCTCGCAAT 786
Db 794 GAGTCTAAGTTCGTAGCGGGCTCAACCTTTTGATGTAAGGGGATTTACCACCACCAAT 853
Qy 787 TCTGTAATCAAAGTG 801
Db 854 ATAGCCTTCAAGTG 868

RESULT 12
LOCUS AB018559
DEFINITION Citrullus lanatus mRNA for wts2L, complete cds.
ACCESSION AB018559
VERSION AB018559.1 GI:6683478
KEYWORDS wts2L.
SOURCE Citrullus lanatus developing seeds cDNA to mRNA.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Rosidae; eurosids I; Cucurbitales; Cucurbitaceae; Citrullus.
REFERENCE
AUTHORS Kim, J., Kang, H., Jun, S., Lee, J. and An, G.
TITLE Cloning of three gibberellin-regulated cDNAs from watermelon during
early seed development: down-regulated one cDNA and up-regulated
two cDNAs
JOURNAL Published Only in DataBase (2000)
AUTHORS 2 (bases 1 to 992)
TITLE Direct Submission
AUTHORS Kim, J., Kang, H., Jun, S., Lee, J. and An, G.
JOURNAL Submitted (12-OCT-1998) Junyul Kim, Pohang University of Science
and Technology (POSTECH), Department of Life Science, Hyoja, Pohang
790-784, Republic of Korea (E-mail:flower4@postech.ac.kr,
Tel:82-562-279-5995)
FEATURES
source Location/Qualifiers
1..992
/organism="Citrullus lanatus"
/db_xref="taxon:3654"
/dev_stage="developing seeds"
/label="synonym:Citrullus vulgaris"
1..992
/gene="wts2L"
65..886
/gene="wts2L"
/codon_start=1
/product="wts2L"
/protein_id="BA89230.1"
/db_xref="GI:BA89230.1"
/translation="MTLNKSPPLRLDGVKVAITGASGIGTSVAVRIFHENGAKVII
ADIDELGGKTIIDELDDVSYIHCDSKEDVSNLVDAAVHRHGKLDIMVSNAGVLDI
SFISILDVKSADYKDLGVNVMGAFWAKHAARVMIPEKNGCILFTSSATNTAGLSS
HPVAASKAGVLGLVRLAALGQHGIRVNCVAFVATGAGSRDPMQAELETMYTT
992
WANKGRVLKADIAALYLASDDANVYSLNLVDGGYSVYVNPMTLTKLMD"
/gene="wts2L"
/label="27 A nucleotides"
BASE COUNT 257 a 253 c 246 g 236 t

Qy 25 AGAAGGCTAGAAAGAAAAGTTGCGCTTATAACAGGAGGAGGAGTGGAAATGGAGAACC 84
Db 95 AGAAGGCTGGAATGAAAAGTTGCGGATCATCACGGAGGCGCAAGCGGATCGGAACAGT 154
Qy 85 ACAGCAAAACTCTCTCCCAACATGAGCAAGTTGCCATTTGCTGTGATGTCACAGATGAA 144
Db 155 CGATGGGAATTTTCCATGAAATGAGGGAAGTAAATCGCGGATATCCCAAGAGAA 214
Qy 145 TTAGTTCACCTCAGTTGTCGAGGCCATTCGCACTTCCCAATTCACCTACATCCACTGTAT 204
Db 215 ATCGGCCAAAATAATCGCGACGAACCTCGCGAT---GACGTAAGCTATATCCACTCGAT 271
Qy 205 GTTACTAATGAAGACGCTGTTAAATAATGCCCTGGACACACAGTTTCAACCTATGAAAA 264
Db 272 GTGTCAAGGAAGACGACGCTCAGCAATCTGTGGACCGCCCGTGCACCGGATGCAAG 331
Qy 265 CTGGACATTATGTTCAAGCAATGCAAGATTTCTGATCCCAACAGGCCCGCATCATAGAC 324
Db 332 CTGGACATCATGTACAGCAACGCGCGGCTTCGACCGTTCTTTAGCGGCATATTTGGAC 391
Qy 325 AACGAAAAGCAGACTTTGAACGCGTTCTCAGTGTAAATGTAACCGGAGTTTTCCTATGC 384
Db 392 GTTACAAAATCTGACTTGGACAAAGCTGTGGCGTGAGCTGTGGGAGCATTTTGGGGA 451
Qy 385 ATGAAGACGCGACGACGCTGTTATCATTCAGACGCGAGTGGCAACATATTTCCACTGCT 444
Db 452 GCAAGCATGCGAGCCAGAGTAATGATACCGAGAAAATGGGTGCAATCTATTCCAAAGC 511
Qy 445 AGTTAAGCTCAACTATGSGGCTGCTCTTCCATGCTATTTGTGTTCAAGCATGCT 504
Db 512 AGCGCAACCAACCAATGCTGCGCTCTCATCACACCATACGCGAGCTTCCAAATGCGCT 571
Qy 505 GTGTTAGCCCTTACTAGGATGTCAGTCGAGTCGAGCAATTTGGCAATTTGGCATTTAGGTTAAT 564
Db 572 GTCTAGGCGTGTGTAGGAACCTAGTCCGAGCTCGGCCAACACGCGGATCAGAGTCAAC 631
Qy 565 TGTTTGTCTCTTTTCGGGCTTCTAGGCTTTAGCAAGAAATTTTCAGGGATTAATAAT 624
Db 632 TGTTGCGCCCTTCTGCTGCGCACCGGATTCAGGG---TCCAGAGACCCGATGCAA 688
Qy 625 GAAGAAGAATTTGAGAATCTATAAACHTTCCGGGAAATTTGAAAGGTCCAAATTTAAT 684
Db 689 GCGAGGCGCTCGAGACCATGGTCACACCTGGGCCAATCTCAAGGTGCTGCTTAAG 748
Qy 685 GTTGAGATGTTGCCAATGCAAGCTTTATCTGGCTAGTGATGAGGCAAAATACGTGAT 744
Db 749 GCGGACGACATAGCCAGGCTCGCTCTCTTTGCTAGCGACGATGCCAATATGTCAGC 808
Qy 745 GGACACAATCTGTTCAATGAGGAGGTTTCAGCGTCTGCAATTC 788
Db 809 GGCCTTAATCTTGTGGTCGATGAGGCTACAGTGTTGTTCAATCC 852

RESULT 13
LOCUS AF352734
DEFINITION Podophyllum peltatum rhizome secoisolariciresinol dehydrogenase
mRNA, partial cds.
ACCESSION AF352734
VERSION AF352734.1 GI:13752455
KEYWORDS Podophyllum peltatum
SOURCE Podophyllum peltatum
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
Berberidaceae; Podophyllum.
REFERENCE
1 (bases 1 to 834)

```

## ORIGIN

Query Match 27.3%; Score 223.2; DB 8; Length 992;  
 Best Local Similarity 57.6%; Pred. No. 3.2e-46;  
 Matches 440; Conservative 0; Mismatches 318; Indels 6; Gaps 2;

2;

AUTHORS Xia,Z.Q., Costa,M.A., Pelissier,H.C., Davin,L.B. and Lewis,N.G.  
 TITLE Secoisolaricresinol dehydrogenase purification, cloning, and functional expression. Implications for human health protection  
 J. Biol. Chem. 276 (16), 12614-12623 (2001)  
 MEDLINE 21201084  
 PUBMED 11278426  
 REFERENCE 2 (bases 1 to 834)  
 AUTHORS Xia,Z.Q., Costa,M.A., Davin,L.B. and Lewis,N.G.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-FEB-2001) Institute of Biological Chemistry, Washington State University, 461 Clark Hall, Pullman, WA 99164-6340, USA

FEATURES  
 source  
 1..834  
 /organism="Podophyllum peltatum"  
 /db\_xref="taxon:35933"  
 1..>834  
 /note="SDH\_Pp7; NAD-dependent; catalyzes enantiospecific conversion of (-)-secoisolaricresinol into (-)-matairesinol"  
 /codon\_start=1  
 /product="thizome secoisolaricresinol dehydrogenase"  
 /protein\_id="AAK38664.1"  
 /db\_xref="GI:13752456"  
 /translation="MGSTSPDSSTNRLOQVIAITGAGGIGETAKLFVRYGAKVV IADIADRGKVCNNIGSPDVISFVHCDTKDEDVRNEVDTTIAKHGIDIMFGNVG LSTTPYSILEAGNEFKRMDINVFYGAFLVAKHAARVMIPAKKGSIVFTASISFTAG EGVSHVYATKHAVLGLTSLCTELGEYGRVNCVSPVIVASPLITDVGVDSSRVEE LAHQANLKGILLRAEDVADAVAYLAGDESXVSGNLVIDGGYTRTPAPPTALKHG LA"

BASE COUNT 233 a 174 c 214 g 213 t  
 ORIGIN

Query Match 25.2%; Score 206.4; DB 8; Length 834;  
 Best Local Similarity 56.4%; Pred. No. 6.4e-42;  
 Matches 447; Conservative 0; Mismatches 336; Indels 9; Gaps 3;

QY 544 CAATTGGCATTAGGTTAATTGTTGTCTCTCTGGGCTTCTTACGGCTTTAGCAAG 603  
 Db 559 GAGTATGGGATCCGAGTAACTGTGTATCTCTCTATATTTGTCATCCCATTTGTTGACC 618  
 QY 604 AATTTTCAGGATTAATAAATGAAGAAGAAATTGACAAATGTAATAAATTTGCGGAAAT 663  
 Db 619 GATGTGTTTGGAGTGGATTCTAGTAGGG---TTGAGGAATTGGCAGCATCAAGTGCAGAAC 675  
 QY 664 TTGAAAGGTCCTCAAAATTTAATGTTGAGGATGTTGCCAATGCAGCTCTTTATCTGGCTAGT 723  
 Db 676 CTCAAAGGACATTTGCTCAGGCTGAGGATGTCGCGATGTCAGTCGCTATTTGGCAGGG 735  
 QY 724 GATGAGCAAAATACGTGAGTGACACAATCTGTTCAATGATGAGGGTTCAGCGTCTGC 783  
 Db 736 GATGATCTCAAGTATCTGACGGCTCTGAATCTTGTTCATCGTGGGGCTACACCAGAACT 795  
 QY 784 AATTCGTGAATC 795  
 Db 796 AATCCGGCTTTC 807

RESULT 14  
 AY099603  
 LOCUS  
 DEFINITION Arabidopsis thaliana short chain alcohol dehydrogenase, putative (Atlg52340) mRNA, complete cds.  
 ACCESSION AY099603  
 VERSION AY099603.1 GI:20466273  
 KEYWORDS FLI\_CDNA.  
 SOURCE Arabidopsis thaliana.  
 ORGANISM Arabidopsis thaliana

REFERENCE  
 AUTHORS Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H., Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A., Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K., Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K., Hayashizaki,Y. and Shinozaki,K.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-APR-2002) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA  
 COMMENT e-mail for correspondence: arab@sequence.stanford.edu

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-Length cDNA': Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Nguyen,M., Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K., Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P., Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.

Nguyen,M. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W. (SSP/Stanford) contributed equally to this work as PIs.

FEATURES  
 source  
 1..1176  
 /organism="Arabidopsis thaliana"  
 /db\_xref="taxon:3702"  
 /chromosome="1"  
 /clone="RAFL09-66-J21 (R24721)"  
 /note="this clone is in pBluescript  
 ecotype: Columbia"  
 1..1176  
 /gene="Atlg52340"

gene



CDS	122.. 979	/gene="Atlg52340"	/codon_start=1	/product="short chain alcohol dehydrogenase, putative"	
		/protein_id="AAW20454.1"	/db_xref="GI:20466274"	/translation="MSTNTSSSYSLPSORLLGKVALITGGATGIGESIVRLFHKHG	
				AKVYDLDQDLGGEVCKSLRSGESKTAFFIHGDRVEDDISNAVDFVKNFIFKHG	
				LINNAAGLGPAPCDIRNYLSLSEFEMTFDYNKGAFLSKHAAARWIPDKGSLYGLDI	
				VGVGVGVGHSHSVGSKHAYLGLTRSAALGQHGIRVNCVSPYATKIALAHLPFE	
				EREDFAVGFERNFAANANLKGVELTVDVANAVLEFLASDSDRYISDNLMDGGFTC	
				TNSFVYFR	
BASE COUNT	314 a	191 c	307 g	364 t	
ORIGIN					
Query Match	24.6%	Score 201.8;	DB 8;	Length 1176;	
Best Local Similarity	56.2%	Pred. No. 9.3e-41;			
Matches	455;	Conservative 0;	Mismatches 327;	Indels 27;	Gaps 3;
QY	27	AAGGCTAGAGGAAGTTGCCCTTATACAGGAGGAGCCAGTGGAAATTCGAGAAACCC	86		
DB	169	AAGCTTTTGGTAAAGTGGCATGATCACTGGAGGAGCCACAGGATAGTGAGAGCAT	228		
QY	87	AGCAAACTCTTCCCAACATGAGCCAAAGTTGCCATTGCTGATCTCCAAAGTGAAT	146		
DB	229	TGTTCGCTCTGTCCACAGCAGCGTCCCAAGTCTGATTTGATCTGCAAGATGATCT	288		
QY	147	AGGTCACTAGTTGTGAGGCCATTCGGCACTTCCCAATTCACACCTA-----CAT	194		
DB	289	CGGAGGTGAGGTGTGAAAGTCTGCTTCGTGGTGTGATCCAGGAGACGCGTTTTCAT	348		
QY	195	CCACTGTGATGTACTAATGAAGAGCGGTGTAAAAATGCCGTGGACACACAGTTTCAAC	254		
DB	349	CCATGGCGATGTAGAGTGGAAAGATGACATAGCAATGCGGTGACTTTCAGTCAAAA	408		
QY	255	CTATGGAACACTGGACATTAATGTCAGCAATGCGAGGAATTTCTGATCCCAAGCCCGC	314		
DB	409	TTTTGGACGCTTGATATCTATCAATGAGGATTAATGTGAGACACCGTCCCGTGA	468		
QY	315	CATCATAGACAGAAACGACACTTTGAACGCGTCTCAGTGTAAATGTAACCGGAT	374		
DB	469	TATTCGTAATATATGTTGAGTGGTTCGAGATGACCTTTGATGTGATGTGAAAGGAGC	528		
QY	375	TTTCTATGATGAAGCAGCAGCAGCTGTATGATTCAGCAGCCAGTGGCAACATAAT	434		
DB	529	TTTTCTAAGCATGAACATGCAGCTCGTGTATGATGATGATGATGATGATGATGATGAT	588		
QY	435	TTTCACTGCTAGTTTAAGCTCAACTATGGTGGTGGTGTCTTCATGATGATGATGATGAT	494		
DB	589	TTTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	648		
QY	495	AAAGCATGCTGTAGGCTTACTAGCAATCTGCGAGTGGAGTGGAGTGGAGTGGAGTGGAG	554		
DB	649	CAAGCATGCTGTAGGCTTACTAGCAATCTGCGAGTGGAGTGGAGTGGAGTGGAGTGGAG	708		
QY	555	TAGGGTAAATTTGTTCT	608		
DB	709	AGGTGGAATCTGTTTCT	768		
QY	609	TTCAGGGAATTAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGAT	659		
DB	769	GGAGGAAGAAAGAGGAGGATGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	828		
QY	660	AAATTTGAAGGTCACAAATTTATGTTGAGGATGTTGCAATGAGTGTGTTTATCTGCG	719		
DB	829	GAATCTAAAGGGGTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	888		
QY	720	TAGTGATGAGCAAAATACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	779		
DB	889	TAGGAT	948		
QY	780	CTGCAATCTGTAATCAAGTCTTCCAT	808		





Db 42387 GGAGGTGAGGTGTTAAAGTCTGCTTCGGTGGTCCAAAGAGACGGCTTTTTCATC 42446  
 QY 196 CACTGTGATGTTTACTTAATGAGACGGTGTAAATAATGCGGTGCACACACAGTTTCAACC 255  
 Db 42447 CATGGCGATGTTAGAGTGAAGATGACATTAGCAATGCGGTTGACTTTGAGTCAAAAAAT 42506  
 QY 256 TATGGAATACTGGACATTTATGTTCCAGCAATGACAGGAATTTCTGATCCCAACAGGCCCGC 315  
 Db 42507 TTTGGAGCGCTTGATATACATTATCAACATGACAGGATATGAGGACCGCTGCCCTGAT 42566  
 QY 316 ATCATAGACAAGCAAGACGCTTGAACCGGTTCTCAGTGAATTAATTAACCGGAGTT 375  
 Db 42567 ATTCGTAATTAATGTTGAGTGTGAGATGACCTTTGATGTGAATGTGAAGGAGCT 42626  
 QY 376 TTCCTATGATGAAGCAGCAGCAGCTGTATGATTCAGACGCGAGTGGCAACATAAT 435  
 Db 42627 TTTCTAAGCATGAACATGACGCTCGTGAATGATACCGGAGAGAAAGGTCGATAGIT 42686  
 QY 436 TCCACTGCTAGTTTAAGCTCAACTATGGTGGTGGTCTTCCATGCTATPTGTGTTC 495  
 Db 42687 TCCTTATGATGTGGAGGTGTTGGGAGGCGTTGGTCCCAATCTTATGTGTGCTC 42746  
 QY 496 AAGCATGCTGTGTAGCCCTTACTAGGAATCTGGCAGTCCGAGCTCGACAAATTTGGCATT 555  
 Db 42747 AAGCATGCTGTGTAGGCTTGTAGCTAGGAGTGTTCGACGCGAGCTTGGACAGCAGGATA 42806  
 QY 556 AGSGTTAATGTTGTCTCTCTTCGGGCTTCCTAC-----GGCTTTAGGCAAGAAATTT 609  
 Db 42807 CGTGTGAAGTGTGTTTCGCTTACGCGTTGCAACTAACTCGCTTGGCTCATTTGCCG 42866  
 QY 610 TCAGGATTAAATAAGAAAGAAATTTGAGAAATGTAATAAATTTGC-----GGGA 660  
 Db 42867 GAGGAAGAAAGACGAGGATGCATTGTTGTTTTCAGGAATTTGCTGCTCCAAACCG 42926  
 QY 661 AATTTGAAGTCCAAATTTAATGTTGAGGATGTTCCCAATGACGCTCTTAICTGCT 720  
 Db 42927 AATCTAAAGGGGTGGAACTGACGGTTGATGATGATGACGAGCTGCTGTTTGGCT 42986  
 QY 721 AGTGATGAGGCAAAATACGTTAGTGGACACAAATCTGTTCAATGATGAGGCTTCAGCGTC 780  
 Db 42987 AGCGATGACTCCGGTACATAAGCGGAGATAATTTGATGATGATGAGGATTCACCTGC 43046  
 QY 781 TGCAATTCGTAAATCAAGTGTTCCAAT 808  
 Db 43047 ACTAACCACCTCCTTAAGTCTTCAGAT 43074

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2003, 09:18:54 ; Search time 15 Seconds  
(without alignments)  
535.497 Million cell updates/sec

Title: US-09-673-918A-2

Perfect score: 1393

Sequence: 1 MQLRFAFARLEKGVALLTG.....IDGFGVNCVSIKVFQYPS 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/1/1aa/5A-COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/1aa/5B-COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/1aa/6A-COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/1aa/6B-COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS-COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	588.5	42.2	337	1 US-08-440-856A-3	Sequence 3, Appli
2	582.5	41.8	333	1 US-08-440-856A-4	Sequence 4, Appli
3	500	35.9	359	1 US-08-440-856A-8	Sequence 8, Appli
4	369	26.5	256	4 US-09-504-358-14	Sequence 14, Appl
5	369	26.5	256	4 US-09-504-358-14	Sequence 14, Appl
6	349.5	25.1	262	4 US-09-363-189B-6	Sequence 6, Appli
7	341.5	24.5	261	4 US-09-468-738A-29	Sequence 29, Appl
8	341.5	24.5	261	4 US-09-940-019-29	Sequence 29, Appl
9	331.5	23.8	267	4 US-09-134-001C-5042	Sequence 5042, Ap
10	324.5	23.3	257	4 US-09-134-001C-3562	Sequence 3562, Ap
11	320	23.0	255	4 US-08-815-225-4	Sequence 4, Appli
12	317.5	22.8	274	4 US-09-134-001C-4431	Sequence 4431, Ap
13	315	22.6	271	2 US-07-637-865-2	Sequence 2, Appli
14	312.5	22.4	244	1 US-08-375-962B-13	Sequence 13, Appl
15	312.5	22.4	244	2 US-08-562-114B-13	Sequence 13, Appl
16	312.5	22.4	244	4 US-08-729-594A-13	Sequence 13, Appl
17	312.5	22.4	244	4 US-08-937-993-13	Sequence 13, Appl
18	304.5	21.9	249	4 US-09-134-001C-4825	Sequence 4825, Ap
19	301.5	21.6	246	3 US-09-238-481-2	Sequence 2, Appli
20	301.5	21.6	246	4 US-09-572-810A-2	Sequence 2, Appli
21	294.5	21.1	252	3 US-08-822-322-8	Sequence 8, Appli
22	294.5	21.1	252	4 US-09-466-109-8	Sequence 8, Appli
23	293	21.0	256	1 US-08-594-808B-7	Sequence 7, Appli
24	272	19.5	283	4 US-09-367-012-1	Sequence 1, Appli
25	272	19.5	283	4 US-09-777-157A-1	Sequence 1, Appli
26	265.5	19.1	243	4 US-09-239-052-2	Sequence 2, Appli
27	265.5	19.1	292	4 US-09-468-738A-2	Sequence 2, Appli

28	265.5	19.1	292	4 US-09-940-019-2	Sequence 2, Appli
29	265.5	19.1	296	4 US-09-468-738A-23	Sequence 23, Appl
30	265.5	19.1	296	4 US-09-940-019-23	Sequence 23, Appl
31	264	19.0	258	4 US-09-504-358-12	Sequence 12, Appl
32	264	19.0	258	4 US-09-954-314-12	Sequence 12, Appl
33	264	19.0	283	4 US-09-134-001C-5346	Sequence 5346, Ap
34	262	18.8	315	3 US-08-793-035-9	Sequence 9, Appli
35	262	18.8	315	3 US-08-793-035-10	Sequence 10, Appl
36	256.5	18.4	247	1 US-08-241-766-13	Sequence 13, Appl
37	256	18.4	261	4 US-08-815-225-2	Sequence 2, Appli
38	256	18.4	261	4 US-08-815-225-3	Sequence 3, Appli
39	256	18.4	261	4 US-09-347-878-50	Sequence 50, Appl
40	256	18.4	290	4 US-09-134-001C-4339	Sequence 4339, Ap
41	254.5	18.3	244	2 US-09-090-567-2	Sequence 2, Appli
42	252.5	18.1	248	4 US-09-385-028-11	Sequence 11, Appl
43	251.5	18.1	303	3 US-09-002-298-1	Sequence 1, Appli
44	250.5	18.0	244	1 US-08-762-129-3	Sequence 3, Appli
45	248	17.8	263	4 US-09-134-001C-4512	Sequence 4512, Ap

ALIGNMENTS

RESULT 1  
US-08-440-856A-3  
; Sequence 3, Application US/08440856A  
; Patent No. 5750873  
; GENERAL INFORMATION:  
; APPLICANT: DELLAPORTA, STEPHEN L.  
; TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING  
; TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 PENNSYLVANIA AVE. N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/440,856A  
; FILING DATE: 15-MAY-1995  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MILLMAN, ROBERT A.  
; REGISTRATION NUMBER: 36,217  
; REFERENCE/DOCKET NUMBER: 05463-20001.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1517  
; TELEFAX: (202) 887-0763  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 337 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-440-856A-3

Query Match 42.2%; Score 588.5; DB 1; Length 337;  
Best Local Similarity 42.2%; Pred. No. 2.6e-55;  
Matches 122; Conservative 50; Mismatches 88; Indels 29; Gaps 4;

QY	5	TAFARLEKGVALLTGAGSIGETTAKLFSQHCARKVAIADVQDELGHVSVEALGTSNSTY	64
DB	47	TPMKRLDGKVAIVTGAGGIGEAIVRLFAKHGARVVIADIDDAAGEALASALGPQVS-F	105
QY	65	IHCDDVTNEDGVKNVAVDNTVSTY-GKLDIMFSNAGI---SDPNRPRIDNEKADFVRLSV	120

```

Db 106 VRCDVSEDDVRAVDWALSRHGRDLVYCNAGVLGRQTRAARSILSFDAAEFRVLRV 165
QY 121 NVTGVFLCMKHAARVMIPARSNGIITASLSSTMGSGSHAYCGSKHAVLALTRNLAVEL 180
Db 166 NALGAALCMKHAARAPRAGSIVSVASVAVGLGLGPHAYTASHKHAIVGLTKNAACEL 225
QY 181 GQFGRVNCVCLSPFGLPTA-----LGKKFSGIKNEEEFENV 216
Db 226 RAHGVVNCVSPFGVATPMLINAWROGHDDATADADRLDLDLVTPSDQVEKMEEVV 285
QY 217 NFAGNLKCPKPNVEDVANAALYLASDEAKYVSGHNLFDGGFSVCNSVI 265
Db 286 RGLATLKGPTLRPDIAEAVFLASDEARIYSGHNLVVDGGVTTSRNLI 334

RESULT 2
US-08-440-856A-4
; Sequence 4, Application US/08440856A
; Patent No. 5750873
; GENERAL INFORMATION:
; APPLICANT: DELLAPORTA, STEPHEN L.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING
; TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVE. N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/440,856A
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLMAN, ROBERT A.
; REGISTRATION NUMBER: 36,217
; REFERENCE/DOCKET NUMBER: 05463-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1517
; TELEFAX: (202) 887-0763
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-440-856A-4

Query Match 41.8%; Score 582.5; DB 1; Length 333;
Best Local Similarity 42.2%; Pred. No. 1.1e-54;
Matches 121; Conservative 50; Mismatches 89; Indels 27; Gaps 4;

QY 5 TAFARRLEGKVALITGGASGIGETTAKLFSQHGAKVAIADVDQLGHSVVEAIGTSNSTY 64
Db 46 TPKPKRLEGKVAIVTGGARGIGEAIVRLFKHGAKVVIADIDDAAGEALAAALG-PHVG 104
QY 65 IHCDDVTNEDGVKNADVNTVSTYCKLDIMFSNAGI---SDPNRPRIIDNEKADFERVLSN 121
Db 105 VRCDVSEEDVERAVERAVARYGRDLVLCNNAGVLGRQTRAARKSILSPDAGEFDRVLVN 164
QY 122 VTGVFLCMKHAARVMIPARSNGIITASLSSTMGSGSHAYCGSKHAVLALTRNLAVELG 181
Db 165 ALGAALCMKHAALAMTORRAGSIISVASVAVGLGLGPHAYTASHKHAIVGLTKNAACELG 224

```

```

QY 182 QGIRVNCVCLSPFGLPTAL-----GKKFS-----GIKNEEEFENVINF 218
Db 225 AHGIRVNCISPGFVATPMLINAWROGHDDATADADLDLDTAVPSDQVEKMEEVVRG 284
QY 219 AGNLKGPKNVEDVANAALYLASDEAKYVSGHNLFDGGFSVCNSVI 265
Db 285 LATLKATGLRPDIAEAAFLASDDSRYSIGHNLVVDGGVTTSRNLI 331

RESULT 3
US-08-440-856A-8
; Sequence 8, Application US/08440856A
; Patent No. 5750873
; GENERAL INFORMATION:
; APPLICANT: DELLAPORTA, STEPHEN L.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING
; TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVE. N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/440,856A
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLMAN, ROBERT A.
; REGISTRATION NUMBER: 36,217
; REFERENCE/DOCKET NUMBER: 05463-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1517
; TELEFAX: (202) 887-0763
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-440-856A-8

Query Match 35.9%; Score 500; DB 1; Length 359;
Best Local Similarity 38.8%; Pred. No. 1.1e-45;
Matches 112; Conservative 40; Mismatches 107; Indels 30; Gaps 5;

QY 5 TAFARRLEGKVALITGGASGIGETTAKLFSQHGAKVAIADVDQLGHSVVEAIGTSNSTY 64
Db 50 TPKPKRLEGKVAIVTGGARGIGEAIVRLFKHGAKVVIADIDDAAGEALAAALG-PX-V 108
QY 65 IHCDDVTNEDGVKNADVNTVSTYCKLDIMFSNAGI---SDPNRPRIIDNEKADFERVLSV 120
Db 109 VRCDVSEVDVRAVAVXXAXXXGRDLVXCNNAGVLGRQTRAAXSILSPDAXEFDRLRV 168
QY 121 NVTGVFLCMKHAARVMIPARSNGIITASLSSTMGSGSHAYCGSKHAVLALTRNLAVEL 180
Db 169 NALGAALCMKHAAXAMXXRRAGSIXSVASVAVGLGLGPHAYTASHKHAIVGLTKNAACEL 228
QY 181 GQFGRVNCVCLSPFGLPTA-----LGKKFSGIKNEEEFENV 216
Db 229 XH-GXRVNCXSPFGVATPMLINAWROGHDXXTADXXDXDLDDXXVPSDQVEKMEEVV 287
QY 217 NFAGNLKCPKPNVEDVANAALYLASDEAKYVSGHNLFDGGFSVCNSVI 265

```

Db 288 RGLATLKGXTLRPDIAEAXLFASDXXRYISGHNLVVDGGVTSRNLI 336

## RESULT 4

US-09-504-358-14

; Sequence 14, Application US/09504358

; Patent No. 6365376

; GENERAL INFORMATION:

; APPLICANT: Rouviere, Pierre E.

; APPLICANT: Brzostowicz, Patricia C.

; TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIATES

; FILE REFERENCE: BC1001 US NA

; CURRENT APPLICATION NUMBER: US/09/504,358

; CURRENT FILING DATE: 2000-02-15

; EARLIER APPLICATION NUMBER: 60/120,702

; EARLIER FILING DATE: 1999-February-19

; NUMBER OF SEQ ID NOS: 49

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 14

; LENGTH: 256

; TYPE: PRT

; ORGANISM: Brevibacterium sp HCU

US-09-504-358-14

Query Match 26.5%; Score 369; DB 4; Length 256;

Best Local Similarity 35.7%; Pred. No. 9.1e-32;

Matches 91; Conservative 55; Mismatches 101; Indels 8; Gaps 5;

QY 10 RLEGKVALITGGASGIGETTAKLFSQHGAQVAIADYQDELGHSVVEAIGTSN--STYIHC 67

Db 3 RLGKQVAVITGGAGMGRIQSELYASEGAQVAVVDVNEQEGRATADAIRASGGVANYWKL 62

QY 68 DVTNEDGVKNVAVDNTVSTYCKLDMFNSNAGISDPNRPRIIDNEKADFERSLVSNVTGVFL 127

Db 63 DVSDESEVEIVWSDIAKRGAINVLNNAGVTGADKPTHEIDER-DLDLVLSVDVKGVEF 121

QY 128 CMKHAARVMIPARSNGNIISTASLSSTMGSSHAYCGSKHAVLALTRNLAVELQGFGRV 187

Db 122 MTKHCIPYFKQAGGAIVNFASITGLVGSQELTPYHAAGAVVVALTKQDAVTYGPSNIRV 181

QY 188 NCLSPGLPTALGKFGSGIKNEEFENFVINFAGNLKGPKNV---EDVANAALYLASDEA 244

Db 182 NAVAPGTILTPLVKEL-GSRGPDGLDGYTKLMG-AKHPLGRVGTPEEVAATLFLASEEA 239

QY 245 KYVSGHNLFDGGFS 259

Db 240 SFITGAVLPVDGGYT 254

## RESULT 5

US-09-954-314-14

; Sequence 14, Application US/09954314

; Patent No. 6465224

; GENERAL INFORMATION:

; APPLICANT: Rouviere, Pierre E.

; APPLICANT: Brzostowicz, Patricia C.

; TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIATES

; FILE REFERENCE: BC1001 US NA

; CURRENT APPLICATION NUMBER: US/09/954,314

; CURRENT FILING DATE: 2001-09-17

; PRIOR APPLICATION NUMBER: 60/120,702

; PRIOR FILING DATE: 1999-February-19

; NUMBER OF SEQ ID NOS: 49

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 14

; LENGTH: 256

; TYPE: PRT

; ORGANISM: Brevibacterium sp HCU

US-09-954-314-14

Query Match

Best Local Similarity 26.5%; Score 369; DB 4; Length 256;

Matches 91; Conservative 55; Mismatches 101; Indels 8; Gaps 5;

```
; Patent No. 6312933
; GENERAL INFORMATION:
; APPLICANT: Kimoto, No. 6312933ihiro
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Mitsuhashi, Kazuya
; TITLE OF INVENTION: NOVEL CARBONYL REDUCTASE, METHOD FOR PRODUCING SAID ENZYME, DNA
; TITLE OF INVENTION: ENCODING SAID ENZYME, AND METHOD FOR PRODUCING ALCOHOL USING SAID
; TITLE OF INVENTION: ENZYME
; FILE REFERENCE: 06501-050001
; CURRENT APPLICATION NUMBER: US/09/468,738A
; PRIOR FILING DATE: 1999-12-21
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: JP 1998-171160
; PRIOR APPLICATION NUMBER: JP 1998-363130
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0, reformatted using WordPerfect 5.1
; SEQ ID NO 29
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Bacillus subtilis
; ORGANISM: Bacillus subtilis
US-09-468-738A-29

Query Match      24.5%; Score 341.5; DB 4; Length 261;
Best Local Similarity 34.4%; Pred. No. 8.8e-29;
Matches 93; Conservative 50; Mismatches 98; Indels 29; Gaps 10;

QY 11 LEGKVALITGGASGIGETTAKLFSQHCAKVAI---ADVQD--ELGHSVVEAIGTSNSTYI 65
DB 5 LKGKVVAITGAASGLGKAMAIREFGKEQAKVINYYSNKQDPNEVKEEVIKAGG--EAVVV 62
QY 66 HCDVTNEDGVKNVADNTVSTYGLDIMEFSGNAGISDPNRPRIIDNEK--ADFERVLSVNT 123
DB 63 QGDVTKEDVKNIVQTAKEFGFLDINNAGLENP---VPSHEMPLKDWKVGICTNLT 118
QY 124 GVFLCMKHAARVMPAR-SGNIISTASLSSTMGSGSHAYCGSKHAYLALTRNLAVELGQ 182
DB 119 GAFGSRRAIKFYVENDIKGNVNMSSVHEVIPWPLFVHYAASKGKIKLMTETLALEYAP 178
QY 183 FGIRVNCLSPPGLPTAL-GKKFSGIKNEEFENVINFAGNLKGPKNFVEDVANAALYLAS 241
DB 179 KGI RVNNGIGAINTPINAEKFPADPKADVESMIPM-GYIGEP----EETAAVAAMLAS 233
QY 242 DEAKYVSGHNLFDGGSVCNSVIKVFQYP 271
DB 234 KEASYVTGITLFDGGM-----QYP 254

RESULT 9
US-09-134-001C-5042
; Sequence 5042, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GFC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5042
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5042

Query Match      23.8%; Score 331.5; DB 4; Length 267;
Best Local Similarity 32.1%; Pred. No. 1.1e-27;
Matches 88; Conservative 55; Mismatches 96; Indels 35; Gaps 9;

QY 11 LEGKVALITGGASGIGETTAKLFSQHCAKVAIADVDDELGHSHVVEAIGTSNSTY----- 64
DB 9 LENKVVLTGAATGIGAKIAENFGKAKVVI--NYRSDRHSEIEEIKQTAKFGGQTLA 67
QY 65 IHCVDVTNEDGVKNVADNTVSTYGLDIMEFSGNAGI---SDPNRPRIIDNEKADFERVLSVNV 122
DB 68 VOGDVSEEDIKRIETIINHFGTLDIINNAGFENSIPHEMSID---DWQKVIDINL 123
QY 123 TGVFLCMKHAARVMPA-RSGNIISTASLSSTMGSGSHAYCGSKHAYLALTRNLAVELG 181
DB 124 TGAIVGSRRETINQFLKENKGTIINISVHDITPWPVHYAASKGGLKLMETMSMEYA 183
QY 182 QFGRVNCLSPPGLPTALGK-KFSGIKNEEFENVINFAGNLKGPK---FNVEDVANAAL 237
DB 184 QYGRINNISPGAIVTEHTKEKSDPTRETIKMI-----PAREIGNAQDVANAVL 235
QY 238 YLASDEAKYVSGHNLFDGGSVCNSVIKVFQYP 271
DB 236 FLSSDLASYHTGTLTYVDGG-----MMNYP 260
```







Db 63 PASIESVLEKIRAEFGVDILVNNAGITRDNLMLRMKDEWND---IETNLSSVFLSK 119  
QY 131 HAARVMIPARSGNIISTASLSSTMGSGSHAYCGSKHAVLALTRNLAVELGQFGRVNCCL 190  
Db 120 AVNRAMMKRHRGRIITIGSVVGTMGNGGOANYAAAKAGLIGFSKSLAREVASRGITVNVV 179  
QY 191 SPFGLPTALGKFKSGIKNEEFENVNFAGNLKGPKNVEDVANAALYLASDEAKYVSGH 250  
Db 180 APGFIETDMTRALSDQORAGILAQVP--AGRLGG---AQEIANAVAFLASDEAAVITGE 233  
QY 251 NLFIDGG 257  
Db 234 TLHVNGG.240

RESULT 15  
US-08-562-1148-13  
; Sequence 13, Application US/08562114B  
; Patent No. 5972646  
; GENERAL INFORMATION:  
; APPLICANT: ERIKSSON ET AL.  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE WHICH CODES FOR A  
; TITLE OF INVENTION: 32 KDA PROTEIN HAVING 11-CIS RETINOL HYDROGENASE  
; TITLE OF INVENTION: ACTIVITY, AND WHICH ASSOCIATES WITH P63, A  
; TITLE OF INVENTION: PORTION OF A RETINOL BINDING PROTEIN RECEPTOR  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felfe & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York City  
; STATE: New York  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage  
; COMPUTER: IBM  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect 5.1 and ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/562,114B  
; FILING DATE: 22-No. 5972646ember-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/375,962  
; FILING DATE: 20-January-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kohli, Vineet  
; REGISTRATION NUMBER: 37,003  
; REFERENCE/DOCKET NUMBER: LUD 5372.2 CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 688-9200  
; TELEFAX: (212) 838-3884  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 244 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: E.coli 3-oxoacyl[acyl carrier protein]reductase  
; NAME/KEY:  
US-08-562-1148-13

Query Match 22.4%; Score 312.5; DB 2; Length 244;  
Best Local Similarity 31.6%; Pred. No. 1.1e-25;  
Matches 78; Conservative 51; Mismatches 107; Indels 11; Gaps 5;  
QY 12 EGKVALITGASGIGETTAKLFSQHGAKVAIAVDQDELGHVSVEAIGTSNSTVIHCDVTN 71  
Db 4 EGKIALVTGASRGIGRAIAETLAARGKV-IGTATSENGAQAISDYLGANGKGLMLNVTD 62  
QY 72 EDGVKNVADNTVSTYKGLDIFMSNAGISDPN-RPRIIDNEKADFRLVSNVTGVLCKMK 130

Db 63 PASIESVLEKIRAEFGVDILVNNAGITRDNLMLRMKDEWND---IETNLSSVFLSK 119  
QY 131 HAARVMIPARSGNIISTASLSSTMGSGSHAYCGSKHAVLALTRNLAVELGQFGRVNCCL 190  
Db 120 AVNRAMMKRHRGRIITIGSVVGTMGNGGOANYAAAKAGLIGFSKSLAREVASRGITVNVV 179  
QY 191 SPFGLPTALGKFKSGIKNEEFENVNFAGNLKGPKNVEDVANAALYLASDEAKYVSGH 250  
Db 180 APGFIETDMTRALSDQORAGILAQVP--AGRLGG---AQEIANAVAFLASDEAAVITGE 233  
QY 251 NLFIDGG 257  
Db 234 TLHVNGG 240  
Search completed: February 14, 2003, 09:21:39  
Job time : 17 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2003, 09:18:54 ; Search time 33 Seconds  
(without alignments)  
1704.572 Million cell updates/sec

Title: US-09-673-918A-2

Perfect score: 1393  
Sequence: 1 MQLTAFARLEKGVKALITG.....IDGFSVNSVIKVFQYVDS 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1364	97.9	277	10	Q94KL7
2	734	52.7	267	10	P93697
3	700.5	50.3	271	10	Q94G09
4	693.5	49.8	271	10	Q94G10
5	685	49.2	284	10	O50038
6	677	48.6	284	10	O82465
7	673	48.3	277	10	Q9SBD8
8	670	48.1	284	10	Q9SBD6
9	668	48.0	277	10	Q9SBD7
10	665.5	47.8	273	10	Q9SBM0
11	658.5	47.3	285	10	Q9SBM0
12	652.5	46.8	278	10	Q94KL8
13	625.5	44.9	303	10	Q9SCU0
14	597.5	42.9	283	10	Q9ZRL7
15	593.5	42.6	259	10	Q93Y47
16	590	42.4	300	10	Q9LW35

17	590	42.4	331	10	P93795	P93795 tripsacum d
18	579.5	41.6	259	10	Q9LEG3	Q9LEG3 lycopersico
19	575.5	41.3	296	10	Q9LS70	Q9LS70 arabidopsis
20	574.5	41.2	257	10	Q94K41	Q94K41 arabidopsis
21	567	40.7	257	10	O80713	O80713 arabidopsis
22	565	40.6	306	10	Q9LW34	Q9LW34 arabidopsis
23	560	40.2	306	10	Q94L84	Q94L84 arabidopsis
24	559	40.1	251	10	O40133	O40133 lycopersico
25	549.5	39.4	258	10	O80714	O80714 arabidopsis
26	543.5	39.0	268	10	O9SQF9	O9SQF9 pisum sativ
27	542.5	38.9	268	10	O9SQJ3	O9SQJ3 pisum sativ
28	534.5	38.4	294	10	Q94FD0	Q94FD0 zea mays su
29	534.5	38.4	294	10	Q94FC9	Q94FC9 zea mays su
30	534.5	38.4	294	10	O93W09	O93W09 zea mays su
31	534.5	38.4	294	10	O93VX5	O93VX5 zea mays su
32	534.5	38.4	294	10	O93VX4	O93VX4 zea mays su
33	533.5	38.3	296	10	O94FC8	O94FC8 zea mays su
34	533.5	38.3	296	10	O94FC7	O94FC7 zea mays su
35	533.5	38.3	296	10	O93W76	O93W76 zea mays su
36	533.5	38.3	296	10	O93W75	O93W75 zea mays su
37	508.5	36.5	281	10	O41345	O41345 silene prat
38	503.5	36.1	281	10	O42407	O42407 silene lati
39	499.5	35.9	234	10	O40590	O40590 nicotiana t
40	490.5	35.2	272	10	O9M1K9	O9M1K9 arabidopsis
41	487.5	35.0	272	10	O24452	O24452 arabidopsis
42	471.5	33.8	256	10	O9SQJ2	O9SQJ2 pisum sativ
43	468.5	33.6	329	10	P93796	P93796 tripsacum d
44	432	31.0	253	16	O91376	O91376 pseudomonas
45	431	30.9	160	10	Q8S9B7	Q8S9B7 ipomoea tri

#### ALIGNMENTS

#### RESULT 1

Q94KL7  
ID Q94KL7 PRELIMINARY; PRT; 277 AA.  
AC Q94KL7;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE Stem secoisolariciresinol dehydrogenase (Fragment).  
OS Forsythia intermedia (Border forsythia).  
OC Eukaryota; Viridiplantae; Streptophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Lamiales; Oleaceae; Forsythia.  
OX NCBI\_TaxID=55183;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21201084; PubMed=11278426;  
RA Xia Z.O., Costa M.A., Pelissier H.C., Davin L.B., Lewis N.G.;  
RT "Secoisolariciresinol Dehydrogenase Purification, Cloning, and  
Functional Expression. IMPLICATIONS FOR HUMAN HEALTH PROTECTION.";  
RL J. Biol. Chem. 276:12614-12623(2001).  
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
(SDR) FAMILY.  
DR EMBL: AF352735; AAK38665.1; -.  
DR InterPro: IPR002198; ADH\_short.  
DR Pfam: PF00106; adh\_short.1.  
DR PROSITE: PS00061; ADH\_SHORT; UNKNOWN\_1.  
KW Oxidoreductase.  
FT NON\_TER 277  
SQ SEQUENCE 277 AA: 29256 MW; 98885C210CAFE2EB CRC64;

Query Match 97.9%; Score 1364; DB 10; Length 277;  
Best Local Similarity 98.2%; Pred. No. 2.9e-95;  
Matches 267; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 QURTAFAARLEKGVKALITGASIGETTAKLFSOHGAKVAIADVDQELGSHVVEAIGTSN 61

Db 6 QVLTATARLEKGVKALITGASIGETTAKLFSOHGAKVAIADVDQELGSHVVEAIGTSN 65

QY 62 STYIHCDVTNEDGVKNAVDNTVSTYTKGLDIFNSAGISDPNRPRIIDNEKADFERYLSVN 121

```
Db 66 STYTHCDVNTEDGVKNAVDNTVSTYKLDIMFNSAGISDPNRPRIIDNEKADFERVESN 125
QY 122 VTGVFLCMKHAARVMIPARSGNIISTASLSMTGGSSHAYCGSKHAYLALTRNLAVELG 181
Db 126 VTGVFLCMKHAARVMIPARSGNIISTASLSMTGGSSHAYCGSKHAYLALTRNLAVELG 185
QY 182 QFGIRVNCLEPFLGPTALGKFKSGIKNEEFENVINFGAGNLKGPKNFVEDVANALYLAS 241
Db 186 QFGIRVNCLEPFLGPTALGKFKSGIKNEEFENVINFGAGNLKGPKNFVEDVANALYLAS 245
QY 242 DEAKYVSGHNLFDGGFVSCNSVIKVFQYPS 273
Db 246 DEAKYVSGHNLFDGGFVSCNSVIKVFQYPS 277

RESULT 2
ID P93697 PRELIMINARY; PRT; 267 AA.
AC P93697;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CPRL12 protein.
OS Vigna unguiculata (Cowpea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
OX NCBI_TaxID=3917;
RN [1]
RP SEQUENCE FROM N.A.
RA Iuchi S., Yamaguchi-Shinozaki K., Urao T., Shinozaki K.;
RT "Characterization of two cDNAs for novel drought-inducible genes in
RT the highly drought-tolerant cowpea.";
RL J. Plant Res. 109:415-424 (1996).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DB EMBL; D88121; BAA13541.1; -.
DR HSSP; P19992; LHDC.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase.
SQ SEQUENCE 267 AA; 28484 MW; 8C4A48A2E9F41B64 CRC64;

Query Match 52.7%; Score 734; DB 10; Length 267;
Best Local Similarity 56.3%; Pred. No. 1e-47;
Matches 147; Conservative 41; Mismatches 67; Indels 6; Gaps 2;

QY 9 RLREGKVALITGGASIGETAKLFSQHAKVAIADVQDELGHVVEAIGTSYTHCD 68
Db 12 KRLGKVAIITGGASIGEGATKLFSGHARVAVADIQDDVGLSLCNEL--KSAIYVHCD 69
QY 69 VTNEDGVKNAVDNTVSTYKLDIMFNSAGISDPNRPRIIDNEKADFERVLSVNTGVFLC 128
Db 70 VTRKEDIKCVDTAVSKFGKLDIMFNAGTGDGFKKSIDNTKSDFERVLSVNLVGPFLG 129
QY 129 MKHAARVMIPARSGNIISTASLSMTGGSSHAYCGSKHAYLALTRNLAVELGFGIRVN 188
Db 130 TKHAARVMIPARSGNIISTASVAGICGGGATHAYTSSKHALVGLTKTAVELGQFGIRVN 189
QY 189 CLSPFLGPTALGKFKSGIKNEEFENVINFGAGNLKGPKNFVEDVANALYLASDEAKYVS 248
Db 190 CVSPFAIVTPLNKNYLDIEGVRKTYMNLKGYVPPV---NDVAEAAALYLASDESKEVS 245
QY 249 GHNLFIDGGFVSCNSVIKVFQ 269
Db 246 SHNLVDIGGLINSVGFPMPE 266

RESULT 3
Q94G09
```

```
ID Q94G09 PRELIMINARY; PRT; 271 AA.
AC Q94G09;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Sex determination protein..
GN CUG..
OS Cucumis sativus (Cucumber).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
OX NCBI_TaxID=3659;
RN [1]
RP SEQUENCE FROM N.A.
RA Sun J.-Q., Li X.-G.;
RT "Gene cloning and expression of CSG in cucumber.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DB EMBL; AF286651; AAK83036.1; -.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase.
SQ SEQUENCE 271 AA; 28997 MW; 7CD6ED7F52AD6224 CRC64;

Query Match 50.3%; Score 700.5; DB 10; Length 271;
Best Local Similarity 53.7%; Pred. No. 3.6e-45;
Matches 146; Conservative 44; Mismatches 75; Indels 7; Gaps 4;

QY 1 MOLRTAFARLEKGVALITGGASIGETAKLFSQHAKVAIADVQDELGHVVEAIGTS 60
Db 3 IQLLPATARLEKGVAVITGGASIGEGATKLFPGHAKVAVIADQHLGOTLCKDLGOS 62
QY 61 NSTYHCDVNTEDGVKNAVDNTVSTYKLDIMFNSAGI--SDPNRPRIIDNEKADFERVL 118
Db 63 SSVFVHCDVTKEKDVETAVDTAVTSKYKLDIMLNAGVFEESPFD-FLKDDPLTFQVW 121
QY 119 SVNTGVFLCMKHAARVMIPARSGNIISTASLSMTGGSSHAYCGSKHAYLALTRNLAV 178
Db 122 NVNLVGAFLGKTHAARVMKPAAGSVITVTASICSVIGGIGTHAYTSSKHVGLMLRNAV 181
QY 179 ELGOGIRVNCLEPFLGPTALGKFKSGIKNEEFENVINFGAGNLK-GPKFNVEDVANAAL 237
Db 182 DLGRYGRVNCVSPNVVPTGMRKLFKVRDGGEP---SFYWSLKNGLDILREEDVGEAVV 238
QY 238 YLASDEAKYVSGHNLFDGGFVSCNSVIKVFQ 269
Db 239 YLGDESKVSGNLIVDGGFTVYNQALCSFR 270

RESULT 4
ID Q94G10 PRELIMINARY; PRT; 271 AA.
AC Q94G10;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE CTA.
GN CTA.
OS Cucumis sativus (Cucumber).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
OX NCBI_TaxID=3659;
RN [1]
RP SEQUENCE FROM N.A.
RA Sun J.-Q., Li X.-G., Zhang X.-S.;
RT "Gene cloning and expression of CTA in cucumber.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DB EMBL; AF286650; AAK83035.1; -.
RN [1]
```



Qy 248 SGHNLFDGGSVCNSVIK 267  
 Db 251 SGLNLLIDGGFTTTNIAFQV 270

RESULT 7  
 Q9SBD8 PRELIMINARY; PRT; 277 AA.

AC Q9SBD8: 2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Short-chain alcohol dehydrogenase.  
 GN SSP.  
 OS Ipomoea trifida.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.  
 OX NCBI\_TaxID=35884;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Koyama Y., Kadota N., Sassa H., Kakeda K., Tsuchiya T., Ishimizu T.,  
 RA Kondo K., Norioka S.;  
 RT "S-luciferin specific stigma protein (SSP) from sporophytic self-  
 RT incompatible plant, Ipomoea trifida, is a member of short-chain  
 RT alcohol dehydrogenase family";  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
 CC (SDR) FAMILY.  
 DR EMBL; AF072448; AAC35341.1; -;  
 DR HSP; P19992; LHDC.  
 DR InterPro; IPR002198; ADH\_short.  
 DR Pfam; PF00106; adh\_short; 1.  
 DR PRINTS; PR00080; SDRFAMILY.  
 KW Oxidoreductase.  
 SQ SEQUENCE 277 AA; 29312 MW; 7A75BFF12A73D9AE CRC64;

Query Match 48.3%; Score 673; DB 10; Length 277;  
 Best Local Similarity 51.7%; Pred. No. 4.5e-43;  
 Matches 134; Conservative 42; Mismatches 83; Indels 0; Gaps 0;

Qy 9 RLREGKVALITGASGIGETTAKLFSOHGAKVAIADVDELGHGSVVEAIGTSNITYHCD 68  
 Db 12 KRLEGKVAIITGAANGIGATTARLFAHQGCKVIAIDDKNGHGSVAEEIGPEYALFHCD 71  
 Qy 69 VTNEGDKNAVDNTVSYKGLDIFMSNAGISDPNRPRIIDNEKADFVRLSVNVTGVFLC 128  
 Db 72 VRIESHVQHAVDITVSRKGLDIFMSNAGVAGSRDTSILEASPENINLVFETNVFGAFFC 131  
 Qy 129 MKHAARVNIIPARSGNIISTASLSTMTGGSSHAYCGSKHAVIALTRNLAVELGQFGRVN 188  
 Db 132 AKHAARVNIIPARKGVIFSAASAEVFGITSDTYTASKCAVVGKSLCVMGKYGKIKAN 191  
 Qy 189 CLSPFGLPTALGKKFGSKNEEFENVNFAGNLKPKFNVEDVANAALYLASDEAKYVS 248  
 Db 192 CVSPYVILTKLGSIMPTQDRKLAEEIVAEASNFKGKTLTTEDVAEALYLAGDESKFVS 251  
 Qy 249 GHNLFDGGSVCNSVIK 267  
 Db 252 GLNLLIDGGFTTTNIAFQV 270

RESULT 8  
 Q9SBD6 PRELIMINARY; PRT; 284 AA.

AC Q9SBD6: 2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Short-chain alcohol dehydrogenase.  
 GN SSP.  
 OS Ipomoea trifida.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.  
 OX NCBI\_TaxID=35884;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Koyama Y., Kadota N., Sassa H., Kakeda K., Tsuchiya T., Ishimizu T.,  
 RA Kondo K., Norioka S.;  
 RT "S-luciferin specific stigma protein (SSP) from sporophytic self-  
 RT incompatible plant, Ipomoea trifida, is a member of short-chain  
 RT alcohol dehydrogenase family";  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
 CC (SDR) FAMILY.  
 DR EMBL; AF072450; AAC35343.1; -;  
 DR HSP; P19992; LHDC.  
 DR InterPro; IPR002198; ADH\_short.  
 DR Pfam; PF00106; adh\_short; 1.  
 DR PRINTS; PR00080; SDRFAMILY.  
 KW Oxidoreductase.  
 SQ SEQUENCE 284 AA; 29898 MW; 2B54BFFBAD6BAE4A CRC64;

Query Match 48.1%; Score 670; DB 10; Length 284;  
 Best Local Similarity 51.7%; Pred. No. 7.8e-43;  
 Matches 134; Conservative 41; Mismatches 84; Indels 0; Gaps 0;

Qy 9 RLREGKVALITGASGIGETTAKLFSOHGAKVAIADVDELGHGSVVEAIGTSNITYHCD 68  
 Db 12 KRLEGKVAIITGAANGIGATTARLFAHQGCKVIAIDDKNGHGSVAEEIGPEYALFHCD 71  
 Qy 69 VTNEGDKNAVDNTVSYKGLDIFMSNAGISDPNRPRIIDNEKADFVRLSVNVTGVFLC 128  
 Db 72 VRIESHVQHAVDITVSRKGLDIFMSNAGVAGSRDTSILEASPENINLVFETNVFGAFFC 131  
 Qy 129 MKHAARVNIIPARSGNIISTASLSTMTGGSSHAYCGSKHAVIALTRNLAVELGQFGRVN 188  
 Db 132 AKHAARVNIIPARKGVIFSAASAEVFGITSDTYTASKCAVVGKSLCVMGKYGKIKAN 191  
 Qy 189 CLSPFGLPTALGKKFGSKNEEFENVNFAGNLKPKFNVEDVANAALYLASDEAKYVS 248  
 Db 192 CVSPYVILTKLGSIMPTQDRKLAEEIVAEASNFKGKTLTTEDVAEALYLAGDESKFVS 251  
 Qy 249 GHNLFDGGSVCNSVIK 267  
 Db 252 GLNLLIDGGFTTTNIAFQV 270

RESULT 9  
 Q9SBD7 PRELIMINARY; PRT; 277 AA.

AC Q9SBD7: 2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Short-chain alcohol dehydrogenase (S-locus linked stigma  
 DE protein).  
 GN SSP OR SSP10.  
 OS Ipomoea trifida.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.  
 OX NCBI\_TaxID=35884;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Koyama Y., Kadota N., Sassa H., Kakeda K., Tsuchiya T., Ishimizu T.,  
 RA Kondo K., Norioka S.;  
 RT "S-luciferin specific stigma protein (SSP) from sporophytic self-  
 RT incompatible plant, Ipomoea trifida, is a member of short-chain  
 RT alcohol dehydrogenase family";  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE EMBL/GenBank/DBJ databases.  
 CC (SDR) FAMILY.  
 DR EMBL; AF072450; AAC35343.1; -;  
 DR HSP; P19992; LHDC.  
 DR InterPro; IPR002198; ADH\_short.  
 DR Pfam; PF00106; adh\_short; 1.  
 DR PRINTS; PR00080; SDRFAMILY.  
 KW Oxidoreductase.  
 SQ SEQUENCE FROM N.A.  
 RC STRAIN=H77-2;  
 RA Tsuchiya T., Suwabe K., Watase T., Kagaya Y., Koyama Y.;



```

Qy 8 ARLEGKVALITGASGIGETAKLFSOHGAKVAIADVDLGHSHVVEAL--GTSNST-- 63
Db 15 SORLLGVALITGGIGESIVRFLFHKGAKVCIVLDQDLGGEVCKSLRGESKETAF 74
Qy 64 YIHCDVTNEDGVKNVNDVSTYKGLDIFMSNAGISDPNRIIDNEKADFERVLSVNV 123
Db 75 FIHGDVVEDDISNAVDFAVKNGFTDLILINNAGLCGAPCPDIRNYSLSFEFTFDVNVK 134
Qy 124 GVFLCMKHAARVIMPARSGNIISTASLSTMTGGSSHAYCGSKHAVLALRNLAVALGQF 183
Db 135 GAPLSMKHAARVIMPEKSGISVLSVCSGVGVGPHSVGSKHAVLGLTRSAVAALGOH 194
Qy 184 GIRVNCSPGLTALGKFKSGIKNEEFENVI---NFA---GNLKGPKFNVEDVANA 236
Db 195 GIRVNCSPVAVATKA--LAHLPEERTEDAFCVFRFAANANLKGVELTVDDVANAV 252
Qy 237 LYASDPAKYVSGHNLFDIGGFSVCNSVTKVQ 269
Db 253 LFLASDDSRISGDNLMIDGGFTCTNHSFRVFR 285

RESULT 12
Q94KL8
ID Q94KL8 PRELIMINARY; PRT: 278 AA.
AC Q94KL8;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Rhizome secoisolariciresinol dehydrogenase (Fragment).
OS Podophyllum peltatum (Mayapple).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
OC Berberidaceae; Podophyllum.
OX NCBI_TaxID=35933;
RN [1]
RP SEQUENCE FROM N.A.
EX MEDLINE=21201084; PubMed=11278426;
RA Xia Z.Q., Costa M.A., Pelissier H.C., Davin L.B., Lewis N.G.;
RT "Secoisolariciresinol Dehydrogenase Purification, Cloning, and
RT Functional Expression. IMPLICATIONS FOR HUMAN HEALTH PROTECTION."
RL J. Biol. Chem. 276:12614-12623(2001).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
(SDR) FAMILY.
DR EMBL; AF352734; AAK38664.1; -.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
KW NON_TER
FT 278
SQ SEQUENCE 278 AA; 29253 MW; DB735A376E112375 CRC64;

Query Match 46.8%; Score 652.5; DB 10; Length 278;
Best Local Similarity 52.2%; Pred. No. 1.6e-41;
Matches 133; Conservative 43; Mismatches 76; Indels 3; Gaps 3;

Qy 10 RLECKVALITGASGIGETAKLFSOHGAKVAIADVDLGHSHVVEAIGTSN-STYIHC 68
Db 13 RLQDKVALITGGAGGIGETAKLFRYKGVVIAIDADHCKVCNNIGSPDVISFVHCD 72
Qy 69 VTNEGVKNVNDVSTYKGLDIFMSNAGISDPNRIIDNEKADFERVLSVNVTVGFLC 128
Db 73 VTQKEDVRNLVDVTIAKHGKLDIMFNGVGLSTTPYSILEAGNEDFKRVDINYGAFV 132
Qy 129 MKHAARVIMPARSGNIISTASLSS--TMGGSSHAYCGSKHAVLALRNLAVALGQFIRV 187
Db 133 AKHAARVIMPAKKSIVFTASISSETAGEGVSHVYATKHAVLGLTSLCTELGEYIRV 192
Qy 188 NCLSPFLGPTALGKFKSGIKNEEFENVIINFAGNLKGPKFNVEDVANAALYASDEAKYV 247
Db 193 NCVSPYIVASPLLTDFGV--DSSRVEELAHQAANLKGTLLEAEDVADAVAYLAGDESKYV 251

```

```

Qy 248 SGHNLFTDGGFSVCN 262
Db 252 SGLNLVIDGGYTRTN 266

RESULT 13
Q9SCU0
ID Q9SCU0 PRELIMINARY; PRT: 303 AA.
AC Q9SCU0;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Short-chain alcohol dehydrogenase-like protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Delsen M., Berger C., Cooke R., Greillet F., Laudie M., Mewes H.W.,
RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
(SDR) FAMILY.
DR EMBL; ALI32968; CAB63154.1; -.
DR HSP; P19992; IHDC.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short.1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase.
RN [1]
RP SEQUENCE 303 AA; 32151 MW; DC1BD28DABE38DD6 CRC64;

Query Match 44.9%; Score 625.5; DB 10; Length 303;
Best Local Similarity 47.5%; Pred. No. 2e-39;
Matches 131; Conservative 52; Mismatches 78; Indels 15; Gaps 6;

Qy 5 TAFARRLEGKVALITGASGIGETAKLFSOHGAKVAIADVDLGHSHVVEAIGTSNST- 63
Db 26 TLYPKRLEGKVAITGGAGHIGKATVMLFARHGATVVIADVDNVAGSLAKSLSHKTS 85
Qy 64 ---YIHCDVTNEDGVKNVNDVSTYKGLDIFMSNAGI--SDPNRIIDNEKADFERV 118
Db 86 MVAFISCDVSVEADENLVNVTVARYGRDLILFNAGVLGDQKKHKSILDFDADEFD 145
Qy 119 SVNVTGVFLCMKHAARVIMIP-ARSGNIISTASLSTMTGGSSHAYCGSKHAVLALRN 177
Db 146 RVNRYGVGLGKMGKARAMIKRGKCIISTASVAGVMGMGPAYATSKHAIVGLTKNA 205
Qy 178 VELQOFGIRVNCISPLGLPTAL---GKFKSG--TKNE--BEFENVINFAGNLKGPKF 229
Db 206 CELYKGIKRVNCISPLGCVATSMVLNARNTSGDVEDDDVEEMEERFVRSANLKG 265
Qy 230 EDVANAALYASDEAKYVSGHNLFDIGGFSVCNSVI 265
Db 266 NDIAEAALYASDESKYVNGHNLVVDGGVTTARNCV 301

RESULT 14
Q92R17
ID Q92R17 PRELIMINARY; PRT: 283 AA.
AC Q92R17;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE Putative alcohol dehydrogenase.
GN F4C21.6 OR A4G03140.

```







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2003, 09:18:49 ; Search time 29 Seconds  
(without alignments)  
390.450 Million cell updates/sec

Title: US-09-673-918A-2

Perfect score: 1393

Sequence: 1 MQLRTAFARRLEGGKVALITG.....IDGFSVCNSVIKFOYFDS 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	588.5	42.2	336	1 TS2_MAIZE	P50160 zea mays (m
2	406	29.1	250	1 LINC_PSEPA	P50197 pseudomonas
3	396	28.4	255	1 YWFD_BACSU	P39640 bacillus su
4	381.5	27.4	256	1 Y019_THEMEA	Q56318 thermotoga
5	376	27.0	250	1 LINC_PSEPA	P50198 pseudomonas
6	366	26.3	260	1 YK02_MYCTU	Q10855 mycobacteri
7	363	26.1	246	1 LINC_THEMEA	Q9x248 thermotoga
8	357	25.6	248	1 FABG_CHLPN	Q9z8p2 chlamydia p
9	346	24.8	251	1 Y325_THEMEA	Q9wy90 thermotoga
10	345.5	24.8	261	1 YGCV_ECOLI	P76633 escherichia
11	341.5	24.5	261	1 DHG_BACSU	P12310 bacillus su
12	334	24.0	253	1 Y4NP_RHISN	P55575 rhizobium s
13	332.5	23.9	248	1 FABG_CHLMU	Q9pkf7 chlamydia m
14	332	23.8	263	1 UCFA_SALTY	P37441 salmonella
15	328.5	23.6	256	1 DHG_RHOSH	Q59787 rhodobacter
16	326.5	23.4	247	1 FABG_CHLTR	P38004 chlamydia t
17	320.5	23.0	244	1 FABG_VIBCH	Q9kdh7 vibrio chol
18	320.5	23.0	262	1 YXBG_BACSU	P46331 bacillus su
19	320	23.0	255	1 2BHD_STREX	P19392 streptomyce
20	316.5	22.7	244	1 FABG_ECOLI	P25716 escherichia
21	315.5	22.6	261	1 DHG_BACME	P40288 bacillus me
22	315	22.6	270	1 DHMA_FLAS1	P22441 flavobacter
23	314.5	22.6	261	1 DHG1_BACME	P39482 bacillus me
24	313.5	22.5	241	1 FABG_RICPR	P50941 rickettsia
25	310	22.3	263	1 UCFA_ECO57	Q8xbj4 escherichia
26	310	22.3	263	1 UCFA_ECOLI	P37440 escherichia
27	309.5	22.2	261	1 DHGA_BACME	P10528 bacillus me
28	308.5	22.1	244	1 FABG_SALTY	O85141 salmonella
29	308.5	22.1	261	1 DHG2_BACME	P39483 bacillus me
30	308	22.1	261	1 DHBA_BACSU	P39071 bacillus su
31	307.5	22.1	244	1 FABG_VIBHA	P55336 vibrio harv
32	306.5	22.0	261	1 DHG3_BACME	P39484 bacillus me
33	306.5	22.0	261	1 DHG4_BACME	P39485 bacillus me

ALIGNMENTS

RESULT 1				
TS2_MAIZE				
ID	TS2_MAIZE	STANDARD;	PRT;	336 AA.
AC	P50160;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Sex determination protein tasselseed 2.			
GN	TS2.			
OS	Zea mays (Maize).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;			
OC	Panicoidae; Andropogoneae; Zea.			
OX	NCBI_TaxID=4577;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. W22;			
RX	MEDLINE=93364991; PubMed=8358795;			
RA	Delong A., Calderon-Urrea A., Dellaporta S.L.;			
RT	*Sex determination gene TASSELSEED2 of maize encodes a short-chain			
RT	alcohol dehydrogenase required for stage-specific floral organ			
RT	abortion.*			
RL	Cell 74.757-768(1993).			
CC	-!- FUNCTION: REQUIRED FOR STAGE-SPECIFIC FLORAL ORGAN ABORTION.			
CC	-!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDOUCTASES			
CC	(SDR) FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: L20621; AAC37345.1; --			
DR	HSSP: P19992; IHDC.			
DR	MaizeDB; 56963; --			
DR	InterPro: IPR002198; ADH_short.			
DR	Pfam: PF00106; adh_short; 1.			
DR	PRINTS: PR00080; SDRFAMILY.			
DR	PROSITE: PS00061; ADH_SHORT; 1.			
KW	Oxidoreductase; Developmental protein.			
FT	NP_BIND 59 83 NAD OR NADP (BY SIMILARITY).			
FT	ACT_SITE 207 207 BY SIMILARITY.			
SQ	SEQUENCE 336 AA; 35204 MW; 4E273D6152B0BB99 CRC64;			
Query Match 42.2%; Score 588.5; DB 1; Length 336;				
Best Local Similarity 42.2%; Pred. No. 7.5e-40;				
Matches 122; Conservative 50; Mismatches 88; Indels 29; Gaps 4;				
QY	5	TAFARLEGGKVALITGASIGETTAKLFSQHGAKVAIADVDQLGHSVVEAIGTSNSTY	64	
Db	47	TPMPKRLDGKVAIVTGGARGIGAIIVLRFAGHGARVVIADIDDAAGEALASALGPQVS-F	105	
QY	65	IHCDEVNEDGVKNAVDNTVSTY-GKLDIMPSNAGI---SDPNRPRIIDNEKADFVRLSV	120	



```
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73124; CAA51638.1; -
DR EMBL; 299123; CAB15799.1; -
DR PIR; S39737; S39737.
DR HSSP; P50162; IAE1.
DR Subtilist; BG10628; ywfD.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Hypothetical protein; Oxidoreductase; Complete proteome.
FT NP_BIND 11 33 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 154 154 BY SIMILARITY.
SQ SEQUENCE 255 AA; 27324 MW; 20AA2259BFB88C9B CRC64;

Query Match 28.4%; Score 396; DB 1; Length 255;
Best Local Similarity 36.6%; Pred. No. 1.2e-24;
Matches 94; Conservative 48; Mismatches 99; Indels 16; Gaps 4;

QY 11 LGGKVALITGGASGIGETTAKLFSQHGAKVATADVODELGHVSVEAIGTSNTYIHCQVT 70
DB 5 LDKTKVLTITGGASGIGYAAVQAFLGQQANVVVADIDEAQGEAMVRKENDRLHFVQTDT 64
QY 71 NEDGVKNAVNDVTSTYCKLDIMFSNAGIS--DPNRPRIIDNEKADFERVLSVNVTVGVFLC 128
DB 65 DEACQAHAVESAVHTFGGLDVLINNGIETVAP----IHEMLSDMNKVLQVNLTMGMFLM 120

QY 129 MKHAARVMIPARSGNIISTASLSSTMTGGSSHAYCGSKHAVLALTRNLAVELGQFGIRVN 188
DB 121 SKHALKHLMAAGCKNIINTCSVGGVLVWPDPAYNASKGGVLQLKSMADVAKHQIRVN 180

QY 189 CLSPFGLPTALGKKFSGINKNEEFENV-----INFAGNLKPKFNVEDVANAALYLASD 242
DB 181 CVCPGIDITPLNEKSFLENNEGLEEKKEKAVNPVLLRLGKP----EBIANVMLFLASD 236

QY 243 EAKYVSGHNLFTDGGFSV 259
DB 237 LSSYWTGSAITADGGYT 253

RESULT 4
Y019_THEME
AC Q56318; STANDARD; PRT; 256 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative oxidoreductase TM0019 (EC 1.-.-.-).
GN TM0019.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogae (class); Thermotogales;
OC Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=96125254; PubMed=8550425;
RA Kletzlin A., Adams M.;
RT "Molecular and phylogenetic characterization of pyruvate and 2-
RT ketoisovalerate ferredoxin oxidoreductases from Pyrococcus furiosus
RT and pyruvate ferredoxin oxidoreductase from Thermotoga maritima.";
RL J. Bacteriol. 178:248-257(1996).
```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X85171; CAA59459.1; -
DR EMBL; AE001690; AAD35113.1; -
DR HSSP; O70351; IE6W.
DR TIGR; TM0019; -
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Hypothetical protein; Oxidoreductase; NADP; Complete proteome.
FT NP_BIND 9 33 NADP (BY SIMILARITY).
FT ACT_SITE 153 153 BY SIMILARITY.
FT CONFLICT 130 136 RGGVII -> TRWRSDH (IN REF. 1).
SQ SEQUENCE 256 AA; 28078 MW; D68160BID7980C6B CRC64;

Query Match 27.4%; Score 381.5; DB 1; Length 256;
Best Local Similarity 35.7%; Pred. No. 1.8e-23;
Matches 91; Conservative 54; Mismatches 99; Indels 11; Gaps 5;

QY 11 LGGKVALITGGASGIGETTAKLFSQHGAKVATADVODELGHVSVEAIGTS--NSTYIHC 68
DB 2 LGGKVAVVTGGGGIGAGAAQLFAENGKVKVIAEIDEEAGVEREMLRGLDVTVEKTD 61
QY 69 VTNEDGVKNAVNDVTSTYCKLDIMFSNAGISDPNRPRIIDNEKADFERVLSVNVTVGVFLC 128
DB 62 VADENSVRKNVRKTVIEYGGVDVLVNNAAVMVK--SIFERPLEEWEVIRVNLTPGYIC 119

QY 129 MKHAARVMIPARSGNIISTASLSSTMTGGSSHAYCGSKHAVLALTRNLAVELGQFGIRVN 188
DB 120 SRYCAEEMIKRGGVYIINIASTRAQSEPDTEPYASKGLVALTHSLAVLSRYHVRV 179

QY 189 CLSPFGLPTALGKKFSGINKNEE--EFENYINFGNLKPKFNVEDVANAALYLASDE-AK 245
DB 180 SISPGWIETSEWKKSLRKKPDLRPIDHQHPAGRVGNPL----DIAHLCLVFLADDEKAG 235

QY 246 YVSGHNLFTDGGFSV 260
DB 236 FITGTNFIVDGGMTV 250

RESULT 5
LINX_PSEPA
ID LINX_PSEPA STANDARD; PRT; 250 AA.
AC P50198;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase (EC 1.1.-.-)
GN LINX.
```



Qy	127	LCKWHAARVMPARSNGNIISTASLSSTMGGSSHAYCGSKHAVLALTRNLAVELGOGFGR	189
Dd	:	: : : : : :	:
Dd	119	NVTGVVPYPIKORNGSIVNVSVGIYNPGOTNYAASKAGVIGTKTWAKELAGRNR	178
Qy	187	VNCILSPGLTALCKKFSGIKNEEFENVINFNAGLKGPKNVEDVANAALYLASDEAKY	246
Dd	:	: : : : : :	:
Dd	179	VNAVAPGFETPTEKLUPEKARETALSRIIP--LGRFGKP-----EEVAQVILFLASDESSY	232
Qy	247	VSGHNLFIDGGFSV	260
Dd	:	: : : : : :	:
Dd	233	VTGOVGIDGGLVI	246
 RESULT 8 FABG_CHLPN			
ID	FABG_CHLPN	STANDARD;	PRT; 248 AA.
AC	Q9Z8P2: Q9JOD1:		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	3-oxyacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (3-ketoacyl-		
DE	acyl carrier protein reductase).		
GN	FABG OR CPN0296 OR CP0462.		
OS	Chlamydia pneumoniae (Chlamydophila pneumoniae).		
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.		
OX	NCBI_TaxID=83558;		
[1]	SEQUENCE FROM N.A.		
RP	STRAIN=CWL029;		
RC	MEDLINE=99206606; PubMed=10192388;		
RX	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,		
RA	Olinder L., Grimwood J., Davis R.W., Stephens R.S.;		
RA	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";		
RL	Nat. Genet. 21:385-389(1999).		
RP	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=AB39.		
RX	MEDLINE=20150255; PubMed=10684935;		
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,		
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,		
RA	Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,		
RA	Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L,		
RA	Eisen J., Fraser C.M.;		
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia		
RL	pneumoniae AK39.";		
RL	Nucleic Acids Res. 28:1397-1406(2000).		
[3]	SEQUENCE FROM N.A.		
RC	STRAIN=J138.		
RX	MEDLINE=20330349; PubMed=10871362;		
RA	Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,		
RA	Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;		
RT	"Comparison of whole genome sequences of Chlamydia pneumoniae J138		
RL	from Japan and CWL029 from USA".;		
RL	Nucleic Acids Res. 28:2311-2314(2000).		
CC	-I- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +		
CC	NADP(+) = 3-oxyacyl-[acyl-carrier protein] + NADPH.		
CC	-I- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS		
CC	PATHWAY.		
CC	-I- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES		
CC	(SDR) FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-		
CC	tion between the Swiss Institute of Bioinformatics and the EMBL outstat-		
CC	the European Bioinformatics Institute. There are no restrictions on no-		
CC	use by non-profit institutions as long as its content is in no-		
CC	modified and this statement is not removed. Usage by and for commer-		
CC	entities requires a license agreement (See http://www.isb-sib.ch/anno-		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
EMBL:	AE001614; AAD18445.1; -		
DR	EBL:	AE002207; AAF38299.1; -	





```

RT -K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features."
RL DNA Res. 4:91-113(1997).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U29579; AAA69285.1; ALT_INIT.
DR EMBL; AE000361; AAC75816.1; ALT_INIT.
DR EMBL; D90893; BAA16569.1; ALT_INIT.
DR HSSP; P25529; 1AHH.
DR Ecogene; EG13130; YGOW.
DR InterPro; IPR002198; ADH_short..
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Hypothetical protein; Oxidoreductase; Complete proteome.
FT NP_BIND 22 46 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 166 166 BY SIMILARITY.
SQ SEQUENCE 261 AA; 28042 MW; 2D6B2382768A4F7E CRC64;

Query Match 24.8%; Score 345.5; DB 1; Length 261;
Best Local Similarity 32.3%; Pred. No. 1.4e-20;
Matches 82; Conservative 58; Mismatches 101; Indels 13; Gaps 6;

QY 11 LRGKVALITGGASGETTAKLFSGHAKVAIADVQDELGHS--VVEAIGTSNSTVIHCD 68
DQ 16 LRGKTAIVTGGNSGLGQAFAMALAKAGANIFTPSFVKDNGETKEMIEKGV-EVDFMQVG 74
QY 69 VTNEDGVKNADVNTYSTYKGLDIFSNAGISDPNRIIDNEKADPERVLSVNVTVGLFC 128
DQ 75 ITAEGAPQRIIAACCCERFCTVDILVNNAGICKLN--KVLDGFRADWDPMIDVNLTAAPFL 132
QY 129 MKHAARVMTPARSGNTIISTASLSMTSGGSHAYCGSKHVLALTENLAVELGQFGIRVN 188
DQ 133 STEAAKIMIPQSGKGIINCSLFSYLGQWSPAYSATKHALAGFTKAYCDELGYQNIQVN 192
QY 189 CLSP--FGLPTALGKFKSGIKNEEFENVINPAGNLKPKFNVEDVANAALYLASDEAKY 246
DQ 193 GIAPGYIATDITLATSNETNORVLDHI--PANRWG---DTQDLMGRAVFLASPASNY 246
QY 247 VSGHNLFDIGGFSV 260
DQ 247 VNGHLLVVDGGYIV 260

RESULT 11
DHG_BACSU STANDARD; PRT; 261 AA.
AC P12310; P94430;
DT 01-OCT-1989 (Rel. 12, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glucose 1-dehydrogenase (EC 1.1.1.47).
GN GDH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86168021; PubMed=3082854;
RA Lampel K.A., Uratani B., Chaudhry G.R., Ramaley R.F., Rudikoff S.;
RT "Characterization of the developmentally regulated Bacillus subtilis
RT glucose dehydrogenase gene."
RL J. Bacteriol. 166:238-243(1986).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=97124189; PubMed=8969502;
RA Yanane K., Kumano M., Kurita K.;
RT "The 25 degrees-36 degrees region of the Bacillus subtilis chromosome:
RT determination of the sequence of a 146 kb segment and identification
RT of 113 genes."
RL Microbiology 142:3047-3056(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Berto M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Chol S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Katamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardiniois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowaka A., Seror S.J., Serror P., Shin B.S., Solido B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -!- CATALYTIC ACTIVITY: Beta-D-glucose + NAD(P)(+) = D-glucono-1,5-
CC lactone + NAD(P)H.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- INDUCTION: IT IS INDUCED AT STAGE III OF THE SPORULATION.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M12276; AAA22463.1; -.
DR EMBL; D50453; BAA09024.1; -.
DR EMBL; Z99106; CAB12201.1; -.
DR PIR; S36090; S36090.
DR HSSP; P50162; 1AEL.
DR Subtilist; BG10545; gdh.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Oxidoreductase; NADP; Sporulation; Complete proteome.
FT NP_BIND 11 35 NADP (BY SIMILARITY).
FT ACT_SITE 158 158 BY SIMILARITY.
FT CONFLICT 148 150 EVI -> AF (IN REF. 1).
SQ SEQUENCE 261 AA; 28090 MW; 5894C17DB8F14965 CRC64;

```

```

Query Match          24.5%; Score 341.5; DB 1; Length 261;
Best Local Similarity 34.4%; Pred. No. 2.8e-20;
Matches 93; Conservative 50; Mismatches 98; Indels 29; Gaps 10

QY 11 LEGKVALITGASGIGETTAKLFSGHAKVAI---ADVD--ELGHSVVEAIGTSNSTYI 65
Db 5 LKGKVAITGAASGLGRAMAIRGKQAKVINYYSNKQDPNEVKKEVIKAGG--EAVVV 62

QY 66 HCDVTNEDGVKNADVNTYSTYKGLDTMFSNAGISDPNPRIDNEK--ADPERVLSVNVVT 123
Db 63 QGDVTKEDYKNIQVTAIRKEGFLDTMINNAGLNP----VPSHEMLPKDWKRVIGNLT 118

QY 124 GFVLCMKHARVMIPAR-SCNIIITSLSTSMGCGSSHAYCGSKHAVALT/TRLNVELQG 182
Db 119 CAGFLGSREAIKYPVENDIRKGNVLSVSVHEVTPWLFVHYAASGGKGLKMTETILALEYAP 178

QY 183 FGIRVNCLSFGPLTAL-CKKSGTGKNEEPENVINFAGNLKGPKNVEDVANALYLAS 241
Db 179 KGI RVNIGPGAINTPINAKEKPADKQADVESMTPM-GYICEP----EETAAVAWLAS 233

QY 242 DEAKYVSGHNLFTDGGFSCVNSVIKVFQYP 271
Db 234 KEASYVTGITLEADGGMI-----QYP 254

RESULT 12
ID YAMP_RHISN STANDARD; PRT; 253 AA.
AC P55575;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Putative short-chain type dehydrogenase/reductase YAMP (EC 1.-.-.-).
GN YAMP.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sub pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1].
RP SEQUENCE FROM N.A.
RA MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA "Welt X.; basis of symbiosis between Rhizobium and legumes.";
RT "Molec Biol 34:401(1997)".
FL "Nucleic Acids Res 25:1011-1016(1997)".
CC 1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDS), FAMILY: HIGH, TO BACTERIAL FABG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; AE000085; AAB91779.1;
DR HSSP; P29132; 1DFT.
DR InterPro; IPR002198; ADH short.
DR Pfam; PF00106; adh_short.1.
DR PROSITE; PS00061; ADH_SHORT.1.
KW Hypothetical protein; Oxidoreductase; Plasmid
NP_BIND 10 35 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 159 159 BY SIMILARITY.
FT ACT_SITE 159 159 BY SIMILARITY.
SQ SEQUENCE 253 AA; 25994 MW; B8208A0B0F5F542F CRC64;

Query Match          24.0%; Score 334; DB 1; Length 253;
Best Local Similarity 34.2%; Pred. No. 1.1e-19;
Matches 90; Conservative 55; Mismatches 92; Indels 26; Gaps 10;

QY 11 LEGKVALITGAS--GIGETTAKLFSGHAKVAIDVQ--DELGHSVVEAIGTSNSTYI-- 65

```

```
Best Local Similarity 32.9%; Pred. No. 1.4e-19;
Matches 85; Conservative 52; Mismatches 98; Indels 23; Gaps 7;

QY 11 LECKVALITGGASGIGETAKLFSQHGAKVAIADVOELGSHVVEAIGT---SNSTYIHC 67
Db 5 LVNKAIVTGGSGIGGFGIAKFAEHGANVQIWGINEEAGKSAQDLSDTKSGKVSFALV 64
QY 68 DVTNEDGVKNADVTYTYGKLDIMFSNAGIS-DPNRPRIIDNEKADFERVLSVNTGVF 126
Db 65 DVSKNDMVSAQVQKFLAEYGTIDVVNNAGITRDSLMLRMSEEE---WSVIDTNLGSIY 121
QY 127 LCMKHARVWIPARSNIIISTASLTSMGSGSHAYCGSKHVALTRNLAVELGQFGIR 186
Db 122 NVCSAVIRPMIKARSGAIVNISSIVGLRSGPGQTNFAAKAGIIGFSKALSKEVGSKNIR 181
QY 187 VNCLSPFGLPTALGKFFS-GIKNEEEFENVINPAGNLKG-PRFNV---EDVANAALYLAS 241
Db 182 VNCIARGFDITDMTKLSNLKNEW-----LKGVLGRVGTPEIANAALFLAS 230
QY 242 DEAKYVSGHNLFTDGGFS 259
Db 231 NQSSYITGVLSVDGMA 248

RESULT 14
UCPA_SALTY STANDARD; PRT; 263 AA.
AC P37441;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Oxidoreductase ucpA (EC 1.-.-.-).
GN UCPA OR STM2445 OR STY2682.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SPECIES-S typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [2]
RP SPECIES-S typhimurium; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan K., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davis R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT
```

```
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SCR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008810; RAD21339.1; -
DR PIR; A38121; A38121
DR StyGene; SG10424; ucpa
DR InterPro; IPR002198; ADH_short.
DR PROSITE; PS00061; ADH_SHORT; FALSE_NEG.
KW Oxidoreductase; Complete proteome.
FT CONFLICT 218 L -> M (IN REF. 2).
FT CONFLICT 258 S -> T (IN REF. 2).
FT CONFLICT 263 V -> I (IN REF. 2).
SQ SEQUENCE 263 AA; 27870 MW; 81B26CAD9B4D534E CRC64;

Query Match 23.8%; Score 332; DB 1; Length 263;
Best Local Similarity 33.1%; Pred. No. 1.6e-19;
Matches 87; Conservative 49; Mismatches 113; Indels 14; Gaps 5;

QY 10 RLEGKVALITGGASGIGETAKLFSQHGAKVAIADVOELGSHVVEAIGTSN-STYIHC 68
Db 3 KLTGTALITGASQIGEGIAVFARHGANLILDISDEIKELADELGGGRHCRCTAVKAD 62
QY 69 VTNEDGVKNADVTYTYGKLDIMFSNAGISDPNRPRIIDNEKADFERVLSVNTGVFLC 128
Db 63 VRDFASVQAAVARAKETEGRIDILVNNAGVCRGLGNFLDMSEEDRDFH--IDINIKGVNW 120
QY 129 MKHAARVWIPARSGNIISTASLSSTM-GGSSHAYCGSKHVALTRNLAVELGQFGIRV 187
Db 121 TKAVLPMLKRRKDGRIVMSSVTGDMVADPGETAYALSAAIAGLTSLAVEYAQSGIRV 180
QY 188 NCLSPFGLPTALGKFFSGIKNEEEFENVINPAG-----NLKGPKNFVEDVANAALYLAS 241
Db 181 NAICPGYVTRPNAESTARQSNPDPSVLTEMAKAIPLRLADPL-----EVGLAAFLAS 236
QY 242 DEAKYVSGHNLFTDGGFSVCNV 264
Db 237 DESSYLTTGQNVIDGGSTLPESV 259

RESULT 15
DHSD_RHOSH STANDARD; PRT; 256 AA.
AC Q59787;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Sorbitol dehydrogenase (EC 1.1.1.14) (L-Iditol 2-dehydrogenase)
DE (Polyol dehydrogenase).
GN POLS OR SMOS.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=96032011; PubMed=7551049;
RA Schauder S., Schneider K.-H., Giffhorn F.;
RT "Polyol metabolism of Rhodobacter sphaeroides: biochemical
RT characterization of a short-chain sorbitol dehydrogenase.";
RL Microbiology 141:1857-1863(1995).
CC -!- FUNCTION: CATALYZES THE OXIDATION OF D-GLUCITOL (SORBITOL) TO D-
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2003, 09:18:54 ; Search time 19 Seconds  
(without alignments)  
1381.300 Million cell updates/sec

Title: US-09-673-918A-2  
Perfect score: 1393  
Sequence: 1 MQLRTAFARLEKVALITG.....IDGGFSVCNSVIKVFQYDPS 273  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	734	52.7	267	T11579	probable short cha
2	685	49.2	284	T02257	probable short cha
3	658.5	47.3	285	F95563	hypothetical prote
4	625.5	44.9	303	T46064	short-chain alcoho
5	597.5	42.9	283	H85039	probable alcohol d
6	588.5	42.2	336	A47542	short-chain alcoho
7	575.5	41.3	264	T02174	probable alcohol d
8	567	40.7	257	T02175	probable alcohol d
9	559	40.1	251	T0364	probable short-cha
10	549.5	39.4	258	T02176	probable alcohol d
11	499.5	35.9	234	T03734	short chain alcoho
12	490.5	35.2	272	T47354	alcohol dehydrogen
13	432	31.0	253	F83440	probable short-cha
14	402.5	28.9	258	C70885	probable dehydroge
15	398	28.6	256	A10406	probable dehydroge
16	396	28.4	255	S39737	glucose 1-dehydrog
17	386	27.7	255	G82644	2,5-dichloro-2,5-c
18	381.5	27.4	256	E72427	oxidoreductase, sh
19	370.5	26.6	260	T36846	probable dehydroge
20	366	26.3	260	H70758	probable fabG3 pro
21	363	26.1	246	H72219	3-oxoacyl-(acyl ca
22	363	26.1	254	AC1478	dehydrogenase/redu
23	363	26.1	254	A11177	dehydrogenase/redu
24	361.5	26.0	296	E87260	hypothetical prote
25	357	25.6	248	H86527	oxoacyl (carrier p
26	357	25.6	248	H72096	3-oxoacyl-(acyl-ca
27	356.5	25.6	243	F83838	oxidoreductase (sh
28	356	25.6	246	H84136	3-oxoacyl-(acyl-ca
29	353.5	25.4	286	F91082	probable oxidoredu

ALIGNMENTS

RESULT 1

T11579  
probable short chain alcohol dehydrogenase CPRD12, drought-inducible - cowpea  
C:Species: Vigna unguiculata (cowpea)  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 20-Jun-2000  
C:Accession: T11579  
R:Uchi, S.; Yamaguchi-Shinozaki, K.; Urao, T.; Shinozaki, K.  
J. Plant Res. 109, 415-424, 1996  
A:Title: Characterization of two cDNAs for novel drought-inducible genes in the highl  
A:Reference number: Z17293  
A:Accession: T11579  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-267 <IUC>  
A:Cross-references: EMBL:D88121  
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 52.7%; Score 734; DB 2; Length 267;  
Best-Local Similarity 56.3%; Pred. No. 8.3e-51;  
Matches 147; Conservative 41; Mismatches 67; Indels 6; Gaps 2;

QY	9	RRLEKVALITGGASGIGETTAKLFSQHGKVAIADVDQLGHSVVEAIGTSNITYHCD 68
DB	12	KRLGKRAIITGGASGIGETARLFSQGHVAVADQDDVGLSLCNEL--KSAIVVHCD 69
QY	69	VNEDGVKNVADNTVSTYKLDIMFSNAGISDPNRRPIIDNEKADFERVLSVNVGTGFLC 128
DB	70	VTKEDIERKVDTAIVSKFGKLDIMFNAGTGDEFKKSILDNKSFVRSVNLVGPFLG 129
QY	129	MKHAARVMIPARSGLIISTASTMGSSSHAYCGSKHAVLALTRNLAVELGQFGIRVN 188
DB	130	THAARVMIPARSGLIINASTMGSSSHAYCGSKHAVLALTRNLAVELGQFGIRVN 189
QY	189	CLSPFGLTALGKFKSGIKNEEFENVINFAAGNLKGFKNVEDVANAALYLASDEAKYVS 248
DB	190	CVSPFAIVTPLLKNYENLDEGVKRTYMLKGYVPV---NDVAEALYLASDESKEVS 245
QY	249	GHNLFDGGFSVCNSVIKVFQ 269
DB	246	SHNLVIDGGLINSVNGFPMFE 266

RESULT 2

T02257  
probable short chain alcohol dehydrogenase - common tobacco  
C:Species: Nicotiana tabacum (common tobacco)  
C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 21-Jul-2000  
C:Accession: T02257  
R:Herbers, K.; Moenke, G.; Badur, R.; Sonnewald, U.  
Plant Mol. Biol. 25, 1027-1038, 1995  
A:Title: A simplified procedure for the subtractive cDNA cloning of photoassimilate-r  
A:Reference number: S62698; MUID:96145513; PMID:8555446



A:Cross-references: GB:NC\_001268; NID:g7270184; PIDN:CAB77799.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: AT4g03140  
A:Map position: 4  
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 42.9%; Score 597.5; DB 2; Length 283;  
Best Local Similarity 47.0%; Pred. No. 5.9e-40;  
Matches 126; Conservative 45; Mismatches 86; Indels 11; Gaps 4;

QY 9 RRLGKVALITGGASIGETTAKLFSQHGAIAADVQDELGHSVVEATGTSNTHCD 68  
DB 16 KLEGGKVALITGGASIGTKATKGFISHGAKVIAIDQIQOIGRETQDELGPS-CATFPDC 74  
QY 69 VTNEDGVKNADVNTSTYGLDLMFNSAGISDPNRPRIIDNEKADFERVLSVNVTVGFLC 128  
DB 75 VTRESIDIANADVAVSLHTKLDIMYNAGIPCKTPPSIVDLDLNVFKDINTNVRGMAG 134  
QY 129 MKHAARVMPARGSNIIISTASLSTMTGGSSSHAYCGSKHVALTLRNLAVALGQFGIRVN 188  
DB 135 IKHAARVMPRNSGSIITCAGSVTGMGLAQHTYTSVSKSAVIGIVRSTASELCKHRIRVN 194  
QY 189 CLSPFGLPTA----LGKFSGKIKNEEFENVINFAGNLKPKFNVEDVANAALYLASDE 243  
DB 195 CISPFALTTFVMDERQIYPGV-DSRLIQIVQSTGVNLGVECEPTDVANAAYILASDD 253  
QY 244 AKYVGNHNFIDGFSVCNSVIRKVFQYP 271  
DB 254 SKVYGNHNLVVDGGF---TTVKTLDPP 277

RESULT 6  
A47542  
short-chain alcohol dehydrogenase (EC 1.1.1.-) - maize  
C:Species: Zea mays (maize)  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 31-Mar-2000  
C:Accession: A47542  
R:DeLong, A.; Calderon-Urrea, A.; Dellaporta, S.L.  
Cell 74, 757-768, 1993  
A:Title: Sex determination gene tasselseed2 of maize encodes a short-chain alcohol dehydrogenase  
A:Reference number: A47542; MUID:93364991; PMID:8358795  
A:Accession: A47542  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-336 <DEL>  
A:Cross-references: GB:I20621; NID:g393183; PIDN:AAC37345.1; PID:g393184  
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
C:Keywords: alcohol metabolism; oxidoreductase  
F:56-238/Domain: short-chain alcohol dehydrogenase homology <SADH>

Query Match 42.2%; Score 588.5; DB 2; Length 336;  
Best Local Similarity 42.2%; Pred. No. 3.8e-39;  
Matches 122; Conservative 50; Mismatches 88; Indels 29; Gaps 4;

QY 5 TAFARLEKVALITGGASIGETTAKLFSQHGAIAADVQDELGHSVVEATGTSNSTY 64  
DB 47 TPKPKRLDGKVAITVGARGIGEAIVRLFAKHGARVVIADIDDAAGEALASALGPQVS-F 105  
QY 65 IHCQVTNEDGVKNADVNTSTY-GKLDIMFNSAGI--SDPNRPRIIDNEKADFERVLSV 120  
DB 106 VRCDVSVEDVRRADVRLWSRHGRLDVCVNNAGVLRQTRARSLISFDNAEFDRLRV 165  
QY 121 NVTVGFLCMKHAARVMPARGSNIIISTASLSTMTGGSSSHAYCGSKHVALTLRNLAVAL 180  
DB 166 NALGAALGKHAARMAPRAGSIVSVASVAVLGGGLGHAYTASKHAIVGLTKNAACEL 225  
QY 181 GQFGIRVNCLSPPGLPTA-----LGKFSGKIKNEEFENVNI 216  
DB 226 RAHGVRVNCVSPFGVATPMLINAWROGHDDATADADRLDLDLDTVPSPDQVEKMEEV 285  
QY 217 NFAGNLKPKFNVEDVANAALYLASDEAKYVGNHNFIDGFSVCNSVI 265  
DB 286 RGLATLKGPTRLPRDIAEAVLFUASDEARYISGHNHNLVVDGGVTTSRNLI 334

RESULT 7  
T02174

probable alcohol dehydrogenase [imported] - Arabidopsis thaliana  
N:Alternate names: hypothetical protein F14M4.3

C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 16-Feb-2001  
C:Accession: T02174; E84911  
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;  
submitted to the EMBL Data Library, September 1998  
A:Description: Arabidopsis thaliana chromosome II BAC F14M4 genomic sequence.  
A:Reference number: Z14609  
A:Accession: T02174

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-264 <ROU>

A:Cross-references: EMBL:AC004411; NID:g3522932; PID:g3522952

A:Experimental source: cultivar Columbia

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: E84911

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-264 <STO>

A:Cross-references: GB:AE002093; NID:g3522952; PIDN:AAC34234.1; GSPDB:GN00139

C:Genetics:

A:Gene: F14M4.3; At2g47140

A:Map position: 2

A:Introns: 163/2

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 41.3%; Score 575.5; DB 2; Length 264;

Best Local Similarity 46.9%; Pred. No. 3e-38;

Matches 120; Conservative 49; Mismatches 82; Indels 5; Gaps 4;

QY 7 FARRLEKVALITGGASIGETTAKLFSQHGAIAADVQDELGHSVVEATGTSNSTYIH 66

DB 9 FFSRLDGKIVITGGASIGAESVRLTEHGARVVIQVDELGQNVAVSIGEDKASYH 68

QY 67 CDVTNEDGVKNADVNTSTYGLDLMFNSAGISDPNRPRIIDNEKADFERVLSVNVTVGVF 126

DB 69 CDVTNTEVENAVKFTVEKYGKLDVLFNSAGVIEP-FVSILDNLNLDRTIAINLRGTA 127

QY 127 LCMKHAARVMI-PARGSNIIISTASLSTMTGGSSSHAYCGSKHVALTLRNLAVALGQFGI 185

DB 128 AFIKHAARAMVEKGIKRGSIIVCTTSVAAEIAGTAPHGYTTSKHLLGLIKSASGGLCKYGI 187

QY 186 RVNCLSPFGLPTALGKFKFSIKNEEE-FENVINFAGNLKPKFNVEDVANAALYLASDEA 244

DB 188 RVNGVAPFGVATPL--VCNGFKMEPNVVEQNTSASANLKGIVLKHARHVAAPALFLASDES 245

QY 245 KYVSGHNLFTDGGFSV 260

DB 246 AYVSGQNLAVDGGYSV 261

RESULT 8  
T02175

probable alcohol dehydrogenase At2g47130 [imported] - Arabidopsis thaliana  
N:Alternate names: hypothetical protein F14M4.4

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 16-Feb-2001

C:Accession: T02175; D84911

R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;

submitted to the EMBL Data Library, September 1998

A:Description: Arabidopsis thaliana chromosome II BAC F14M4 genomic sequence.

A:Reference number: Z14609

A:Accession: T02175

```
A>Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-257 <ROU>
A:Cross-references: EMBL:AC004411; NID:g3522932; PID:g3522935
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: D84911
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-257 <STO>
A:Cross-references: GB:AE002093; NID:g3522935; PIDN:AAC34217.1; GSPDB:GN00139
C:Gene: At2g47130; F14M4.4
A:Map position: 2
A:Introns: 5/2
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 40.7%; Score 567; DB 2; Length 257;
Best Local Similarity 45.4%; Pred. No. 1.4e-37;
Matches 118; Conservative 49; Mismatches 81; Indels 12; Gaps 5;

QY 10 RLEGKVALITGAGSIGETAKLFSQHGAKVAITADVDDELGHVSVEAIGTSNITYHCDV 69
DB 5 RLDGKIAITGASGIGAEAVRLFTDQGAKEVIVDFQDELGNVAVSQKDKASFYRCDV 64
QY 70 TNECGKNAVONTYSTYKGLDINFNSNAGISDPNRP-RIIDNEKADFERYLSVNVTVGVFLC 128
DB 65 TNEKEVENAVFTYKYGKLDVLFNSAGVME--QPGSFLDLNLEQFDRTMVNVGAAAF 122
QY 129 MKHAARVMI-PARSGNIISTASLSSTMGSSSHAYCGSKHVALTRNLAVELGQFGLRV 187
DB 123 IKHAARVMEKTRGSICTTSVASEIGGPGPHAYTASKEALLGLVLSACGLCKYGLRV 182
QY 188 NCLSPGLPTALGKFKSGIKNEEFENVINFA--GNLKGPRFNVEDVANAALYLASDEA 244
DB 183 NGVAPYAVATAINS-----RDEETVRVMEYSAAITGLKGVVKARHVAEALFLASDDS 237
QY 245 KYVSGHNLFDGGSVCNSV 264
DB 238 AYVSGQNLAVDGGISVVKPI 257

RESULT 9
T06364
Probable short-chain alcohol dehydrogenase (EC 1.1.1.-) - tomato (fragment)
C:Species: Lycopersicon esculentum (tomato)
C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 11-May-2000
C:Accession: T06364
R:Jacobsen, S.E.; Olszewski, N.E.
Planta 198, 78-86, 1996
A>Title: Gibberellins regulate the abundance of RNAs with sequence similarity to protein
A:Reference number: 215627; MUID:96158488; PMID:8580773
A:Accession: T06364
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-251 <JAC>
A:Cross-references: EMBL:U21801; NID:g717141; PIDN:AAB00109.1; PID:g717142
A:Experimental source: cultivar Moneymaker
C:Gene: GAB3
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; oxidoreductase

Query Match 40.1%; Score 559; DB 2; Length 251;
Best Local Similarity 47.5%; Pred. No. 5.7e-37;
Matches 122; Conservative 46; Mismatches 81; Indels 8; Gaps 5;

QY 12 BGKVALITGASGIGETAKLFSQHGAKVAITADVDDELGHVSVEAIGTSNITYHCDVTN 71

Db 1 ECKVAIITGAASGIGAEASARLFEHGARVVVADIQDELGQKVVDSIGSKASYRHCDVTD 60
QY 72 EDGVKNVNDTVSTYKGLDINFNSNAGISDPNRPRIIDNEKADFERYLSVNVTVGVFLCMKH 131
Db 61 EKQVETVAYAVERYKGLDINFNSVGLF--NFCSVLDMOVLAFDETMANVR-IALAVKH 117
QY 132 AARVMIPAR-SGNIISTASLSSTMGSSSHAYCGSKHVALTRNLAVELGQFGLRVNCL 190
Db 118 AAKVMVDKIRGSICTTSVASEIGGPGPHAYTASKEALLGLVLSACGLCKYGLRVN 177
QY 191 SPFGPLTALGKFKSGIKNEEFENVINFAAGNLKGPKNFVEDVANAALYLASDEAKYVSGH 250
Db 178 SPYGIATPLVTKAYGL-DAALLEAIYGNHKGKVLSTMHVAQSAFLASDESAYTSQ 236
QY 251 NLFIDGGSVCNSVTKV 267
Db 237 NLAVDGGLS---SILKL 250

RESULT 10
T02176
Probable alcohol dehydrogenase At2g47120 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein F14M4.5
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02176; C84911
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;
submitted to the EMBL Data Library, September 1998
A:Description: Arabidopsis thaliana chromosome II BAC F14M4 genomic sequence.
A:Reference number: 214609
A:Accession: T02176
A>Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-258 <ROU>
A:Cross-references: EMBL:AC004411; NID:g3522932; PID:g3522936
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: C84911
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-258 <STO>
A:Cross-references: GB:AE002093; NID:g3522936; PIDN:AAC34218.1; GSPDB:GN00139
C:Gene: At2g47120; F14M4.5
A:Map position: 2
A:Introns: 5/2
A:Note: F14M4.5
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 39.4%; Score 549.5; DB 2; Length 258;
Best Local Similarity 44.5%; Pred. No. 3.3e-36;
Matches 113; Conservative 53; Mismatches 79; Indels 9; Gaps 5;

QY 10 RLEGKVALITGAGSIGETAKLFSQHGAKVAITADVDDELGHVSVEAIGTSNITYHCDV 69
DB 5 RLEGKVALITGAGSIGADAAARLFTDHGAKVIVDVQDELGNVAVLICKDKASFYRCDV 64
QY 70 TNECGKNAVONTYSTYKGLDIMPNSNAGISDPNRPRIIDNEKADFERYLSVNVTVGVFLCM 129
DB 65 TNETEVEDAVKFTYKYGKLDVLFNSAGVLEP-LESFLDFDLERFDRIMAVNVGAAAFI 123
QY 130 KHAARVMI-PARSGNIISTASLSSTMGSSSHAYCGSKHVALTRNLAVELGQFGLRVN 188
DB 124 KHAARVMEKTRGSICTTSVASEIGG--HHGYTASKEALLGLVLSACGLCKYGLRVN 182
QY 189 CLSPGLPTALGK--KFSGIKNEEFENVINFAAGNLKGPKNFVEDVANAALYLASDEAKY 246
DB 183 NGVAPYAVATAINS-----RDEETVRVMEYSAAITGLKGVVKARHVAEALFLASDDS 237
```



Db 183 GVAPYAVATPMTSHDEVTVG-----KQLEDYFDAKILKGMVLKASHVAQAFLASDDSYA 238

Qy 247 VSGHNLFTDGGFSV 260  
: : : : :  
Db 239 ISGQNLAVDGGYTV 252

RESULT 11  
T03734  
short chain alcohol dehydrogenase homolog - common tobacco  
N:Alternate names: TPHP-1 protein  
C:Species: Nicotiana tabacum (common tobacco)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000  
C:Accession: T03734  
R:Kawaoka, A.; Kawamoto, T.; Sekine, M.; Yoshida, K.; Takano, M.; Shimmyo, A.  
Plant J. 6, 87-97, 1994  
A:Title: A cis-acting element and a trans-acting factor involved in the wound-induced ex  
A:Reference number: Z15039; MUID:95004656; PMID:7920706  
A:Accession: T03734  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-234 <RAW>  
A:Cross-references: EMBL:D29976; NID:g531268; PIDN:BAA06241.1; PID:g531269  
A:Experimental source: strain cv. Petite Havana SR-1  
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homolog

Query Match 35.9%; Score 499.5; DB 2; Length 234;  
Best Local Similarity 47.5%; Pred. No. 2.7e-32;  
Matches 112; Conservative 33; Mismatches 82; Indels 9; Gaps 3;

Qy 3 LRTAFARLEKGVALITGGASGIGETTAKLFHQHAKVAIADVQDELGHGSVVEAIGTSNS 62  
: : : : :  
Db 6 LPSPIAKRLEKGVALITGGASGIGAATARLVQHGAKVTTADIQDNLTSLVQEIGNHT 65  
Qy 63 TYIHCDVTNEDGYKNAYDVNTVSTYGKLDMFNSNAGISDPNRPLIIDNEKADFERVLSVV 122  
: : : : :  
Db 66 IFHCNVAVESDVQNVDDATIAKFGLDIMFPQVAGGKSISILDVVRI-IKTVPDNI 124  
Qy 123 TGVPFLCMKHARVMIPARSNG-NIISTASLSSTMGSSSHAYCGSKHAVLALTRNLAVELG 181  
: : : : :  
Db 125 VGAFCAKHAARVMIPPKRFHYHLKLLSRT-----HILHKGNAVLGFSKNIGVELG 177  
Qy 182 QFGIRVNCLSPGLPTALGKKFKSGIKNEEFENVINPAGNLKGFKNVEDVANAAL 237  
: : : : :  
Db 178 KYGIKVCVSPHYISTPLVNLNALGIAERETAERKWFAGAGNLKCALLDEEVEKAVL 233

RESULT 12  
T47354  
alcohol dehydrogenase (ATA1) - Arabidopsis thaliana  
N:Alternate names: protein F18p9.120  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 19-May-2000  
C:Accession: T47354  
R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;  
Mayer, K.F.X.  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z24458  
A:Accession: T47354  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-272 <NYA>  
A:Cross-references: EMBL:AL138654  
A:Experimental source: cultivar Columbia; BAC clone F18p9  
C:Genetics:  
A:Map position: 3  
A:Introns: 132/1  
A>Note: F18p9.120  
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homolog

Query Match 35.2%; Score 490.5; DB 2; Length 272;  
Best Local Similarity 44.6%; Pred. No. 1.7e-31;  
Matches 119; Conservative 46; Mismatches 89; Indels 13; Gaps 7;



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2003, 09:18:49 ; Search time 37 Seconds  
(without alignments)  
983.173 Million cell updates/sec

Title: US-09-673-918A-2  
Perfect score: 1393  
Sequence: 1 MOLRTAFARLEKGVALLTG.....IDGFSVCNSIKVFQYPS 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :			
A_Geneseq_101002.*			
1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*		
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*		
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*		
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*		
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*		
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*		
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*		
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*		
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*		
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*		
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*		
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*		
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*		
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*		
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*		
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*		
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*		
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*		
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*		
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*		
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*		
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*		
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	1393	100.0	273 21	AA1980.DAT.*
2	1388	99.6	272 21	AA1981.DAT.*
3	1364	97.9	277 21	AA1982.DAT.*
4	1233	88.5	276 21	AA1983.DAT.*
5	1233	88.5	276 23	AA1984.DAT.*
6	1223	87.8	273 21	AA1985.DAT.*
7	1175	84.4	277 21	AA1986.DAT.*
8	658.5	47.3	285 23	AA1987.DAT.*
9	654.5	47.0	285 21	AA1988.DAT.*
10	625.5	44.9	303 23	AA1989.DAT.*

11	598.5	43.0	300 21	AA1990.DAT.*
12	598.5	43.0	343 21	AA1991.DAT.*
13	597.5	42.9	283 21	AA1992.DAT.*
14	597.5	42.9	283 23	AA1993.DAT.*
15	588.5	42.2	336 16	AA1994.DAT.*
16	588	42.2	269 23	AA1995.DAT.*
17	578.5	41.5	261 21	AA1996.DAT.*
18	578.5	41.5	276 21	AA1997.DAT.*
19	575.5	41.3	264 23	AA1998.DAT.*
20	575.5	41.3	264 23	AA1999.DAT.*
21	573.5	41.2	260 21	AA2000.DAT.*
22	573.5	41.2	260 23	AA2001.DAT.*
23	569	40.8	280 23	AA2002.DAT.*
24	568.5	40.8	286 21	AA2003.DAT.*
25	567	40.7	257 21	AA2004.DAT.*
26	567	40.7	257 23	AA2005.DAT.*
27	566	40.6	259 23	AA2006.DAT.*
28	565	40.6	257 21	AA2007.DAT.*
29	549.5	39.4	258 23	AA2008.DAT.*
30	492.5	35.4	272 21	AA2009.DAT.*
31	492.5	35.4	275 21	AA2010.DAT.*
32	490.5	35.2	272 23	AA2011.DAT.*
33	487.5	35.0	272 21	AA2012.DAT.*
34	432	31.0	253 22	AA2013.DAT.*
35	403.5	29.0	159 21	AA2014.DAT.*
36	375.5	27.0	146 21	AA2015.DAT.*
37	370	26.6	186 21	AA2016.DAT.*
38	370	26.6	186 21	AA2017.DAT.*
39	370	26.6	186 21	AA2018.DAT.*
40	363	26.1	254 23	AA2019.DAT.*
41	357	25.6	251 20	AA2020.DAT.*
42	349.5	25.1	262 22	AA2021.DAT.*
43	347.5	24.9	262 21	AA2022.DAT.*
44	346.5	24.9	253 23	AA2023.DAT.*
45	343.5	24.7	267 22	AA2024.DAT.*

ALIGNMENTS

RESULT 1	
AA1980.DAT.*	AA1980.DAT.*
ID	AA1980.DAT.*
XX	AA1980.DAT.*
AC	AA1980.DAT.*
XX	AA1980.DAT.*
DT	06-APR-2000 (first entry)
XX	Secoisolaricresinol dehydrogenase protein clone DEHY133.
DE	Secoisolaricresinol dehydrogenase; lignan biosynthetic pathway;
XX	Secoisolaricresinol dehydrogenase; lignan biosynthetic pathway;
KW	lignan; metaresinol; health-protecting lignan; phytoestrogen;
KW	enterolactone; enterodiol; neutriceutical; dietary supplement;
KW	(-)-trachelogenin; antiviral; (-)-podophyllotoxin.
XX	Forsythia intermedia.
OS	Forsythia intermedia.
XX	WO9955846-A1.
PN	WO9955846-A1.
XX	04-NOV-1999.
PD	04-NOV-1999.
XX	23-APR-1999; 99WO-US08975.
PF	23-APR-1999; 99WO-US08975.
XX	24-APR-1998; 98US-0082977.
PR	(UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX	(UNIW ) UNIV WASHINGTON STATE RES FOUND.
PA	(UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX	Xia Z, Costa MA, Davin LB, Lewis NG;
PI	Xia Z, Costa MA, Davin LB, Lewis NG;
XX	WPI; 2000-126356/11.
DR	N-PSDB; AA245730.
XX	New nucleic acid molecule encoding an enzyme involved in lignan
PT	New nucleic acid molecule encoding an enzyme involved in lignan

PT biosynthetic pathway, useful for producing large amounts of lignans -  
 XX Claim 8; Page 46-47; 66pp; English.  
 PS The present sequence represents a secoisolariciresinol dehydrogenase  
 XX protein. The enzyme is involved in the lignan biosynthetic pathway.  
 CC The secoisolariciresinol dehydrogenase proteins have a molecular  
 CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require  
 CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase  
 CC nucleic acids are used for the recombinant expression of the enzymes.  
 CC It is also used to obtain expression or enhanced expression of  
 CC secoisolariciresinol dehydrogenase in plants or to alter lignan  
 CC biosynthesis. The enzyme is used for production of the pharmacologically  
 CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase  
 CC proteins and nucleic acids can be utilized to: elevate or otherwise  
 CC alter the levels of health-protecting lignans, including phytoestrogens  
 CC such as enterolactone and enterodiol, in plant species, including  
 CC vegetables, grains and fruits and to food items incorporating material  
 CC derived from such genetically altered plants; genetically alter plant  
 CC species to provide an abundant, natural supply of lignans useful for  
 CC a variety of purposes, for example as nutraceuticals and dietary  
 CC supplements; to genetically alter living organisms to produce an  
 CC abundant supply of optically pure lignans having desirable biological  
 CC properties, for example (-)-trachelogenin which possesses antiviral  
 CC properties, and (-)-podophyllotoxin.  
 XX Sequence 273 AA;  
 SQ  
 Query Match 100.0%; Score 1393; DB 21; Length 273;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-132;  
 Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MQLTAFARLEGKVALITGGASIGETTAKLFSQHGAKVAIADVDLGHVSVEAIGTS 60  
 DB 1 MQLTAFARLEGKVALITGGASIGETTAKLFSQHGAKVAIADVDLGHVSVEAIGTS 60  
 QY 61 NSTYIHCVDVTNEDGVKNVADNTVSTYKLDIMFNSNAGISDPNRPRIIDNEKADFERVLSV 120  
 DB 61 NSTYIHCVDVTNEDGVKNVADNTVSTYKLDIMFNSNAGISDPNRPRIIDNEKADFERVLSV 120  
 QY 121 NVTGVFLCMKHAARVMIPARSGNIISTASLSMTGGGSHAYCGSKHAVLALTRNLAVEL 180  
 DB 121 NVTGVFLCMKHAARVMIPARSGNIISTASLSMTGGGSHAYCGSKHAVLALTRNLAVEL 180  
 QY 181 GQGIIRVNCILSPFGLPTALGKFKSGIKNEEFENVINFAGNLKGPKNVEDVANAALYLA 240  
 DB 181 GQGIIRVNCILSPFGLPTALGKFKSGIKNEEFENVINFAGNLKGPKNVEDVANAALYLA 240  
 QY 241 SDEAKYVSGHNLFDIDGGFVSCNSVIKVFQYDPS 273  
 DB 241 SDEAKYVSGHNLFDIDGGFVSCNSVIKVFQYDPS 273  
 RESULT 2  
 AAY54420  
 ID AAY54420 standard; Protein; 272 AA.  
 XX AC AAY54420;  
 XX DT 06-APR-2000 (first entry)  
 XX DE Secoisolariciresinol dehydrogenase protein clone DEHW130.  
 XX Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;  
 KW lignan; matairesinol; health-protecting lignan; phytoestrogen;  
 KW enterolactone; enterodiol; nutraceutical; dietary supplement;  
 KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin.  
 XX Forsythia intermedia.  
 OS  
 XX WO995846-A1.  
 XX FN  
 XX PD 04-NOV-1999.

XX 23-APR-1999; 99WO-US08975.  
 PF 24-APR-1998; 98US-0082977.  
 PR (UNIW ) UNIV WASHINGTON STATE RES FOUND.  
 XX Xia Z, Costa MA, Davin LB, Lewis NG;  
 PI WPI: 2000-126356/11.  
 DR N-PSDB; AAZ45743.  
 XX New nucleic acid molecule encoding an enzyme involved in lignan  
 PT biosynthetic pathway, useful for producing large amounts of lignans -  
 XX Example 2: Page 61-63; 66pp; English.  
 PS The present sequence represents a secoisolariciresinol dehydrogenase  
 CC protein. The enzyme is involved in the lignan biosynthetic pathway.  
 CC The secoisolariciresinol dehydrogenase proteins have a molecular  
 CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require  
 CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase  
 CC nucleic acids are used for the recombinant expression of the enzymes.  
 CC It is also used to obtain expression or enhanced expression of  
 CC secoisolariciresinol dehydrogenase in plants or to alter lignan  
 CC biosynthesis. The enzyme is used for production of the pharmacologically  
 CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase  
 CC proteins and nucleic acids can be utilized to: elevate or otherwise  
 CC alter the levels of health-protecting lignans, including phytoestrogens  
 CC such as enterolactone and enterodiol, in plant species, including  
 CC vegetables, grains and fruits and to food items incorporating material  
 CC derived from such genetically altered plants; genetically alter plant  
 CC species to provide an abundant, natural supply of lignans useful for  
 CC a variety of purposes, for example as nutraceuticals and dietary  
 CC supplements; to genetically alter living organisms to produce an  
 CC abundant supply of optically pure lignans having desirable biological  
 CC properties, for example (-)-trachelogenin which possesses antiviral  
 CC properties, and (-)-podophyllotoxin.  
 XX Sequence 272 AA;  
 SQ  
 Query Match 99.6%; Score 1388; DB 21; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-132;  
 Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 QLRTAFAARLEGKVALITGGASIGETTAKLFSQHGAKVAIADVDLGHVSVEAIGTSN 61  
 DB 1 QLRTAFAARLEGKVALITGGASIGETTAKLFSQHGAKVAIADVDLGHVSVEAIGTSN 60  
 QY 62 STYIHCVDVTNEDGVKNVADNTVSTYKLDIMFNSNAGISDPNRPRIIDNEKADFERVLSV 121  
 DB 61 STYIHCVDVTNEDGVKNVADNTVSTYKLDIMFNSNAGISDPNRPRIIDNEKADFERVLSV 120  
 QY 122 VTGVFLCMKHAARVMIPARSGNIISTASLSMTGGGSHAYCGSKHAVLALTRNLAVELG 181  
 DB 121 VTGVFLCMKHAARVMIPARSGNIISTASLSMTGGGSHAYCGSKHAVLALTRNLAVELG 180  
 QY 182 QFGIRVNCILSPFGLPTALGKFKSGIKNEEFENVINFAGNLKGPKNVEDVANAALYLA 241  
 DB 181 QFGIRVNCILSPFGLPTALGKFKSGIKNEEFENVINFAGNLKGPKNVEDVANAALYLA 240  
 QY 242 DEAKYVSGHNLFDIDGGFVSCNSVIKVFQYDPS 273  
 DB 241 DEAKYVSGHNLFDIDGGFVSCNSVIKVFQYDPS 272  
 RESULT 3  
 AAY54413  
 ID AAY54413 standard; Protein; 277 AA.  
 XX AC AAY54413;  
 XX DT 06-APR-2000 (first entry)

XX Secoisolariciresinol dehydrogenase protein clone SMDEHY321.  
DE |||||  
KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;  
KW lignan; matairesinol; health-protecting lignan; phytoestrogen;  
KW enterolactone; enterodiol; neutriceutical; dietary supplement;  
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin.  
XX Forsythia intermedia.  
XX WO9955846-A1.  
XX AC  
XX XX  
XX PD 04-NOV-1999.  
XX PF 23-APR-1999; 99WO-US08975.  
XX XX 24-APR-1998; 98US-0082977.  
XX PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.  
XX PI Xia Z, Costa MA, Davin LB, Lewis NG;  
DR WPI: 2000-126356/11.  
DR N-PSDB; AA245731.  
XX  
PT New nucleic acid molecule encoding an enzyme involved in lignan  
PT biosynthetic pathway, useful for producing large amounts of lignans  
XX  
PS Claim 8; Page 49-50; 66pp; English.  
XX  
CC The present sequence represents a secoisolariciresinol dehydrogenase  
CC protein. The enzyme is involved in the lignan biosynthetic pathway.  
CC The secoisolariciresinol dehydrogenase proteins have a molecular  
CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require  
CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase  
CC nucleic acids are used for the recombinant expression of the enzymes.  
CC It is also used to obtain expression or enhanced expression of  
CC secoisolariciresinol dehydrogenase in plants or to alter lignan  
CC biosynthesis. The enzyme is used for production of the pharmacologically  
CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase  
CC proteins and nucleic acids can be utilized to: elevate or otherwise  
CC alter the levels of health-protecting lignans, including phytoestrogens  
CC such as enterolactone and enterodiol, in plant species, including  
CC vegetables, grains and fruits and to food items incorporating material  
CC derived from such genetically altered plants; genetically alter plant  
CC species to provide an abundant, natural supply of lignans useful for  
CC a variety of purposes, for example as neutriceuticals and dietary  
CC supplements; to genetically alter living organisms to produce an  
CC abundant supply of optically pure lignans having desirable biological  
CC properties, for example (-)-trachelogenin which possesses antiviral  
CC properties, and (-)-podophyllotoxin.  
XX  
SQ Sequence 277 AA;  
Query Match 97.9%; Score 1364; DB 21; Length 277;  
Best Local Similarity 98.2%; Pred. No. 1.6e-129;  
Matches 267; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 QLRTAFARLEGKVALITGASGIGETTAKLFSQHGAKVAIADVDQLGHSVVEAIGTSN 61  
| : | |||||  
DB 6 QVLTATARLEGKVALITGASGIGETTAKLFSQHGAKVAIADVDQLGHSVVEAIGTSN 65  
| : | |||||  
QY 62 STVIHCDVNTNEDGVKNAVDTNVTYCKLDTMFNSNAGISDPNRRPIIDNEKADFERVLSVN 121  
| : | |||||  
DB 66 STVIHCDVNTNEDGVKNAVDTNVTYCKLDTMFNSNAGISDPNRRPIIDNEKADFERVFSVN 125  
| : | |||||  
QY 122 VTGVFLCMKHAARVMTIPARSGNIISTASLSSTMTGGSSHAYCGSKHAVLALTRNLAVELG 181  
| : | |||||  
DB 126 VTGVFLCMKHAARVMTIPARSGNIISTASLSSTMTGGSSHAYCGSKHAVLGLTRNLAVELG 195  
| : | |||||  
QY 182 QGIRVNCILSPFGLPALGKKGKSGIKNEEFENVINFAGNLKGPKNVEDVANALYLAS 241  
| : | |||||  
DB 186 QGIRVNCILSPFGLPALGKKGKSGIKNEEFENVINFAGNLKGPKNVEDVANALYLAS 245  
| : | |||||

QY 242 DEAKYVSGHNLFDGGFSVCNSVIKVFQYPDS 273  
|||||  
DB 246 DEAKYVSGHNLFDGGFSVCNSVIKVFQYPDS 277  
|||||  
RESULT 4  
AAY54416  
ID AAY54416 standard; Protein; 276 AA.  
XX  
XX AAY54416;  
XX AC  
XX XX  
XX DT 06-APR-2000 (first entry)  
XX DE Secoisolariciresinol dehydrogenase protein clone SMDEHY631.  
XX XX  
XX KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;  
KW lignan; matairesinol; health-protecting lignan; phytoestrogen;  
KW enterolactone; enterodiol; neutriceutical; dietary supplement;  
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin.  
XX  
XX OS Forsythia intermedia.  
XX XX  
XX PN WO9955846-A1.  
XX  
XX PD 04-NOV-1999.  
XX XX  
XX PF 23-APR-1999; 99WO-US08975.  
XX XX  
XX PR 24-APR-1998; 98US-0082977.  
XX XX  
XX PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.  
XX PI Xia Z, Costa MA, Davin LB, Lewis NG;  
XX XX  
XX WPI: 2000-126356/11.  
DR N-PSDB; AA245734.  
XX  
PT New nucleic acid molecule encoding an enzyme involved in lignan  
PT biosynthetic pathway, useful for producing large amounts of lignans  
XX  
PS Claim 8; Page 56-57; 66pp; English.  
XX  
CC The present sequence represents a secoisolariciresinol dehydrogenase  
CC protein. The enzyme is involved in the lignan biosynthetic pathway.  
CC The secoisolariciresinol dehydrogenase proteins have a molecular  
CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require  
CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase  
CC nucleic acids are used for the recombinant expression of the enzymes.  
CC It is also used to obtain expression or enhanced expression of  
CC secoisolariciresinol dehydrogenase in plants or to alter lignan  
CC biosynthesis. The enzyme is used for production of the pharmacologically  
CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase  
CC proteins and nucleic acids can be utilized to: elevate or otherwise  
CC alter the levels of health-protecting lignans, including phytoestrogens  
CC such as enterolactone and enterodiol, in plant species, including  
CC vegetables, grains and fruits and to food items incorporating material  
CC derived from such genetically altered plants; genetically alter plant  
CC species to provide an abundant, natural supply of lignans useful for  
CC a variety of purposes, for example as neutriceuticals and dietary  
CC supplements; to genetically alter living organisms to produce an  
CC abundant supply of optically pure lignans having desirable biological  
CC properties, for example (-)-trachelogenin which possesses antiviral  
CC properties, and (-)-podophyllotoxin.  
XX  
SQ Sequence 276 AA;  
Query Match 88.5%; Score 1233; DB 21; Length 276;  
Best Local Similarity 88.2%; Pred. No. 2.8e-116;  
Matches 240; Conservative 14; Mismatches 18; Indels 0; Gaps 0;

QY 2 QLRTAFARLEGKVALITGASGIGETTAKLFSQHGAKVAIADVDQLGHSVVEAIGTSN 61  
| : | |||||  
| : | |||||

Db 5 QLRFAARRLEGGKVALITGGASGVGEVTAQLFSQHGAKVAIADVQDELCHSVVEAIGLSN 64  
 QY 62 STYIHCVDVTNEDGVKNVADNTVSTYKGLDIFMSNAGISDPNRPRIIDNEKADFERVLSVN 121  
 Db 65 STYIHCVDVTNEDGVKNVADNTVSTYKGLDIFMSNAGISDPYKPRVIDNEKADFERVLSVN 124  
 QY 122 VTGVFLCMKHAARVMIPARSGNIISTASUSSTMGSGSHAYCGSKHAYLALTRNLAVELG 181  
 Db 125 VTGVFLCMKHAARVMIPARSGNIISTASUSSTMGSGSHAYCGSKHAYLALTRNLAVELG 184  
 QY 182 QFGIRVNCISPFGLPTALGKKFSGIKNEEFENVPFAGNLKGPKNFVEDVANAALYLAS 241  
 Db 185 QFGIRVNCISPFGLPTALGKKFSGIKNEEFENVPFAGNLKGPKNFVEDVANAALYLAS 244  
 QY 242 DEAKYVSGHNLFDIDGFSVCNSVVKFQYDPS 273  
 Db 245 DEAOYVSGQNLFDIDGFSVCNSAIKLFOYDPS 276

RESULT 5  
 AAO21494  
 ID AAO21494 standard; Protein; 276 AA.  
 XX  
 AC AAO21494;  
 XX  
 DT 15-AUG-2002 (first entry)  
 DE Secoisolariciresinol dehydrogenase protein.  
 XX  
 KW Guaiacyl (G)-lignan; monocotyledon plant; rice; food additive;  
 KW seed-specific transcriptional regulatory region; dehydrogenase; enzyme.  
 XX  
 OS Unidentified.  
 XX  
 PW WO200220548-A1.  
 XX  
 PD 14-MAR-2002.  
 XX  
 PF 04-SEP-2001; 2001WO-0527500.  
 XX  
 PR 07-SEP-2000; 2000US-230632P.  
 XX  
 PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.  
 PA (PHYT-) APPLIED PHYTOLOGICS INC.  
 XX  
 PI Lewis NG, Davin LB, Huang N;  
 XX  
 DR WPI: 2002-425767/45.  
 DR N-PSDB; AAL38432.  
 XX  
 PT Increasing guaiacyl-lignan content in monocotyledon plants, by  
 PT transforming plant with chimeric gene construct having seed-specific  
 PT transcriptional regulator linked to gene encoding protein involved in  
 PT G-lignan formation  
 XX  
 PS Claim 2; Fig 10; 136pp; English.  
 XX  
 CC The invention relates to a method for increasing the guaiacyl (G)-lignan  
 CC content in seeds of a monocotyledon plant, comprising selecting at least  
 CC one protein or enzyme integral to the pathway leading to G-lignan  
 CC formation, stably transforming a monocotyledon plant with chimeric gene  
 CC (CG) constructs having a seed-specific transcriptional regulatory region  
 CC operably linked to a nucleic acid sequence encoding the enzyme. The  
 CC method of the invention is useful for stably transforming a  
 CC monocotyledonous plant (e.g. rice) with CG constructs resulting in  
 CC increased expression of the genes encoded by CG constructs. The G-lignan  
 CC enriched seed composition is useful as a food additive. This  
 CC sequence represents the secoisolariciresinol dehydrogenase protein  
 CC relating to the invention.  
 XX  
 SQ Sequence 276 AA;  
 XX  
 Query Match 88.5%; Score 1233; DB 23; Length 276;

Best Local similarity 88.2%; Pred. No. 2.8e-116;  
 Matches 240; Conservative 14; Mismatches 18; Indels 0; Gaps 0;

QY 2 QLRFAARRLEGGKVALITGGASGVGEVTAQLFSQHGAKVAIADVQDELCHSVVEAIGTSN 61  
 Db 5 QLRFAARRLEGGKVALITGGASGVGEVTAQLFSQHGAKVAIADVQDELCHSVVEAIGLSN 64  
 QY 62 STYIHCVDVTNEDGVKNVADNTVSTYKGLDIFMSNAGISDPNRPRIIDNEKADFERVLSVN 121  
 Db 65 STYIHCVDVTNEDGVKNVADNTVSTYKGLDIFMSNAGISDPYKPRVIDNEKADFERVLSVN 124  
 QY 122 VTGVFLCMKHAARVMIPARSGNIISTASUSSTMGSGSHAYCGSKHAYLALTRNLAVELG 181  
 Db 125 VTGVFLCMKHAARVMIPARSGNIISTASUSSTMGSGSHAYCGSKHAYLALTRNLAVELG 184  
 QY 182 QFGIRVNCISPFGLPTALGKKFSGIKNEEFENVPFAGNLKGPKNFVEDVANAALYLAS 241  
 Db 185 QFGIRVNCISPFGLPTALGKKFSGIKNEEFENVPFAGNLKGPKNFVEDVANAALYLAS 244  
 QY 242 DEAKYVSGHNLFDIDGFSVCNSVVKFQYDPS 273  
 Db 245 DEAOYVSGQNLFDIDGFSVCNSAIKLFOYDPS 276

RESULT 6  
 AAV54414  
 ID AAV54414 standard; Protein; 273 AA.  
 XX  
 AC AAV54414;  
 XX  
 DT 06-APR-2000 (first entry)  
 DE Secoisolariciresinol dehydrogenase protein clone SMDEHY431.  
 XX  
 KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;  
 KW lignan; metairensinol; health-protecting lignan; phytoestrogen;  
 KW enterolactone; enterodiol; neutriceutical; dietary supplement;  
 KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin.  
 XX  
 OS Forsythia intermedia.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 120 /note= "encoded by GTN"  
 FT Misc-difference 122 /note= "encoded by GTN"  
 XX  
 PN WO9955846-A1.  
 XX  
 PD 04-NOV-1999.  
 XX  
 PF 23-APR-1999; 99WO-US08975.  
 XX  
 PR 24-APR-1998; 98US-0082977.  
 XX  
 PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.  
 XX  
 PI Xia Z, Costa MA, Davin LB, Lewis NG;  
 XX  
 DR WPI: 2000-126356/11.  
 DR N-PSDB; AAZ45732.  
 XX  
 PT New nucleic acid molecule encoding an enzyme involved in lignan  
 PT biosynthetic pathway, useful for producing large amounts of lignans  
 XX  
 PS Claim 8; Page 51-52; 66pp; English.  
 XX  
 CC The present sequence represents a secoisolariciresinol dehydrogenase  
 CC protein. The enzyme is involved in the lignan biosynthetic pathway.  
 CC The secoisolariciresinol dehydrogenase proteins have a molecular  
 CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require  
 CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase  
 CC nucleic acids are used for the recombinant expression of the enzymes.



```

XX PF 28-AUG-2001; 2001WO-EP09892.
XX PR 28-AUG-2001; 2001WO-EP09892.
XX PA (FARB ) BAYER AG.
XX PI Tietjen K, Weidner M;
XX DR WPI; 2002-269010/31.
XX PT Identifying plant target proteins for herbicidally active compounds,
XX PT comprising aligning and comparing nucleic acid or amino acid sequences
XX PT from plants with nucleic acid or amino acid sequences from non-plant
XX PT organisms -
XX PS Claim 5; SEQ ID NO 540; 261pp + Sequence Listing; English.
XX CC The invention relates to identifying target proteins
XX CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX CC aligning and comparing nucleic acid or amino acid sequences from plant
XX CC with nucleic acid or amino acid sequences from non-plant organisms using
XX CC suitable search parameters, where plant sequences having an E-value
XX CC greater by a factor of 3 than the E-value of most similar non-plant
XX CC sequences are selected. The polypeptides or nucleic acids encoding them
XX CC are useful for identifying modulators. The identified modulators are
XX CC useful as herbicides.
XX SQ Sequence 285 AA;

Query Match 47.3%; Score 658.5; DB 23; Length 285;
Best Local Similarity 50.9%; Pred No. 4.3e-58;
Matches 139; Conservative 43; Mismatches 78; Indels 13; Gaps 5;

QY 8 ARRLGKVALITGASGIGETRAKLFQSGAKVATADVODELGHVSVEAI--GTSNST-- 63
Db :||| ||||| ||||| :|| :||| ||||| :||| ||||| :||| :|||
QY 15 SQRLLGKVALITGAGIGESIVRLPHKGAQKIVDLODDLGGEVCKSLRGESKETAF 74
Db :||| ||||| ||||| :|| :||| ||||| :||| ||||| :||| :|||
QY 64 YHCDVTNEDGVKNADVNTVSYGKLDIMFNSNAGISDPNRPRIIDNEKADPERVLSVNT 123
Db :||| ||||| ||||| :|| :||| ||||| :||| ||||| :||| :|||
QY 75 FIHGVRVEDDIDNAVDFAVKNFTGLDILINAGLCGAPCPDIRNYSLEFEMTFDYNVK 134
Db :||| ||||| ||||| :|| :||| ||||| :||| ||||| :||| :|||
QY 124 GVFLCMKHAARVIMPARSGNIITASLSMTGGSSHAYCGSKHAVLALTRNLAVELGQF 183
Db :||| ||||| ||||| :|| :||| ||||| :||| ||||| :||| :|||
QY 135 GAFLSMKHAARVIMPEKKGIVSLCSVGGVGVGPHSYGSKHAVLGLTRVAELGQH 194
Db :||| ||||| ||||| :|| :||| ||||| :||| ||||| :||| :|||
QY 184 GIRVNCLSFPLPTALGKFKFSIGIKNEEFENVI---NFA---GNLKGPKFNVEDVANAA 236
Db :||| ||||| ||||| :|| :||| ||||| :||| ||||| :||| :|||
QY 195 GIRVNCVSPYAVATKLA--LAHLPEERTEDEAFVGRFAAANALKGVELTVDDVANAV 252
Db :||| ||||| ||||| :|| :||| ||||| :||| ||||| :||| :|||
QY 237 LYLASDEAKYVSGHNLFDGGFSVCNSVIKVFQ 269
Db :||| ||||| ||||| :|| :||| ||||| :||| ||||| :||| :|||
QY 253 LFLASDDSRVSGDNLMDIDGFTCTNHSFKVFR 285
Db :||| ||||| ||||| :|| :||| ||||| :||| ||||| :||| :|||

RESULT 9
AAG25524
ID AAG25524 standard; Protein; 285 AA.
XX AC AAG25524;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 29625.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX
```

```

PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 21-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.
XX 29-JUN-1999; 99US-0140991.
XX 30-JUN-1999; 99US-0141287.
XX 01-JUL-1999; 99US-0141842.
XX 01-JUL-1999; 99US-0142154.
XX 02-JUL-1999; 99US-0142055.
```



```
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
```

```
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.
```

Query Match 47.0%; Score 654.5; DB 21; Length 285;

Best Local Similarity 50.9%; Pred. No. 1.1e-57; Mismatches 79; Indels 13; Gaps 5;

Matches 139; Conservative 42;

```
QY 8 ARREKQKVALITGGAGIGETTAKLFQHQKAKVAIADVQDELGHVSVEAI--GTSNST-- 63
Db 15 SQRLLGKVALITGGATGIGESIVRLFHKHAKVCIVLDQDLGGEGVCKSLRGESKETAF 74
QY 64 YIHCDVTNEDGVKNVNDVTSTYTKLDMFNSNAGISDPNRPRIIDNEKADPERVLSVNV 123
Db 75 FTHGDVVRVEDDISNAVDFAVKNEFTLDILINNAGLCGAPCPDIRNYSLSEFEFEMTDVNVK 134
QY 124 GVFLCMKHAARVMIPARSGNIISTASLSTMGSGSSHYACGSKHAYLALTRNLAVELQOF 183
Db 135 GAFLSMKHAARVMIPKESKVSIVSLCSVGVGVGPHSVGSKHAYLGLTRSVAAELGOH 194
QY 184 GIRVNCVSPFGLPTALGKKFKSGIKNEEFENVI----NFA----GNLKGPKFNVEDVANAA 236
Db 195 GIRVNCVSPYAVATKLA--LAHLPEERTEDAFVGRNFAANANLKGVELTVGDVANAV 252
QY 237 LYLASDEAKYVSGHNLFDGGFVSNVSIKVFQ 269
Db 253 LFLASDDSRYSIGDNLMDIDGGFTCTNHSFKVPR 285
```

RESULT 10

ABB92624

ID ABB92624 standard; Protein; 303 AA.

XX ABB92624;

XX ABB92624;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 1835.



PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148568.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	14-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	22-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 43.0%; Score 598.5; DB 21; Length 300;		
Best Local Similarity 46.8%; Pred. No. 5.6e-52;		
Matches 126; Conservative 46; Mismatches 86; Indels 11; Gaps 4;		
QY	8	ARRLEKVALITGGASGIGETTAKLFQSHQAKVATADVDELGHSVVEAIGTNSNTHC 67
DB	32	SRKLEKVALITGGASGIGKATAGKFISHGAKVIIADIQPOIGRETEQELGFS-CAYFPC 90
QY	68	DVTNEDGVKNVDNTVSTYKLDIMFSNAGISDPNRPRIIDNEKADFERVLSVNVTVGVFL 127
DB	91	DVTKESDIANAVDFAVSLHTKLDIMYNNAGIPCKTPPSIVDLNLNVDFKVINTVRGVMA 150
QY	128	CMKHAARVMIPARSGNIISTASLSMTGGSGSHAYCGSKHAVLALTRNLAVELGQFGRV 187
DB	151	GIKHAARVMIPRNSGSIICAGSVTGMGLAQHTYSVKSASVIGIVRSTASELCKHRIV 210
QY	188	NCLSPFGLPTA-----LGKKFSGIKNEEFENVINPAGNLKPKFNVEVDANAALYASD 242
DB	211	NCISPFAITTSFVMDMRQIYFGV-DDSRLIQIVQSTGVINGEVCPTDVAANAAYLASD 269
QY	243	EAKYVSGHNLFIIDGGFVSCNSVIKVFQYP 271
DB	270	DSKYVNGHNLVVDGGF-----TTVTKTLDFF 294
RESULT 12		

AAG18619  
ID AAG18619 standard; Protein; 343 AA.  
XX AAG18619;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 20100.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 07-MAY-1999; 99US-0132486.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 24-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 01-JUN-1999; 99US-0136782.  
PR 03-JUN-1999; 99US-0137222.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137502.  
PR 08-JUN-1999; 99US-0137724.  
PR 10-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142054.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 18-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.

PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	16-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 43.0%; Score 598.5; DB 21; Length 343;		
Best Local Similarity 46.8%; Pred. No. 6.8e-52;		
Matches 126; Conservative 46; Mismatches 86; Indels 11; Gaps 4;		
QY	8	ARLEKVALITGAGSIGETAKLFQSHQAKVAIADVODELGHVSVEAIGTNSYTHC 67
Db	75	SRKLEKVALITGAGSIGKATAGKFTSHGAKVIAIQPIQIGRETEQELGPS-CAYFFPC 133
QY	68	DVTNEDGVKNVADNTYSTYKLDIMESNAGISDPNPRIDNEKADFERVLSVNTGTGVL 127
Db	134	DVTKESDIANAVDFVSLTKLDIMYNAGIPCKTPPSIVDLDLNVEDKVININRGVMA 193
QY	128	CMKHAARVMPARSGNIITASLSSTMGGSSHAYCGSHAVLALPRNLAVELGQFGIRV 187
Db	194	GIKHAARVMPRNSGIIICAGSVTGMGLAQHTYSVSKSVIGIVRSTASELCKHRIV 253
QY	188	NCLSPGLPTA-----LGRKFGIKNEEFENVINFAGNLKGFKNVEDVANAALYLASD 242

Db	254	NCISPAITTSFVMDMRQIYPCV-DDSRLIQIVOSTGVLNGEVCPTDVANAAYLASD 312
QY	243	EAKYTSGHNLFDGGFSVCNSVIKVFQYP 271
Db	313	DSKYVNGHNLVVDGGF----TTVKTLDFP 337
RESULT 13		
AAG31552		
ID	AAG31552 standard; Protein; 283 AA.	
XX	AAG31552;	
XX	17-OCT-2000 (first entry)	
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 37910.	
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
XX	termination sequence.	
OS	Arabidopsis thaliana.	
XX	EP1033405-A2.	
XX	06-SEP-2000.	
XX	25-FEB-2000; 2000EP-0301439.	
PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	16-APR-1999;	99US-0128714.
PR	19-APR-1999;	99US-0129845.
PR	21-APR-1999;	99US-0130077.
PR	23-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	28-APR-1999;	99US-0130891.
PR	30-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	04-MAY-1999;	99US-0132407.
PR	05-MAY-1999;	99US-0132484.
PR	06-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	07-MAY-1999;	99US-0132487.
PR	11-MAY-1999;	99US-0132863.
PR	14-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.

PR	16-JUN-1999	99JUS-0139452
PR	16-JUN-1999	99JUS-0139453
PR	17-JUN-1999	99JUS-0139492
PR	18-JUN-1999	99JUS-0139454
PR	18-JUN-1999	99JUS-0139455
PR	18-JUN-1999	99JUS-0139456
PR	18-JUN-1999	99JUS-0139457
PR	18-JUN-1999	99JUS-0139458
PR	18-JUN-1999	99JUS-0139459
PR	18-JUN-1999	99JUS-0139460
PR	18-JUN-1999	99JUS-0139461
PR	18-JUN-1999	99JUS-0139462
PR	18-JUN-1999	99JUS-0139463
PR	18-JUN-1999	99JUS-0139750
PR	18-JUN-1999	99JUS-0139763
PR	21-JUN-1999	99JUS-0139819
PR	22-JUN-1999	99JUS-0139899
PR	23-JUN-1999	99JUS-0140353
PR	23-JUN-1999	99JUS-0140354
PR	24-JUN-1999	99JUS-0140699
PR	28-JUN-1999	99JUS-0140823
PR	29-JUN-1999	99JUS-0140991
PR	30-JUN-1999	99JUS-0141287
PR	01-JUL-1999	99JUS-0141842
PR	01-JUL-1999	99JUS-0142134
PR	02-JUL-1999	99JUS-0142055
PR	06-JUL-1999	99JUS-0142390
PR	08-JUL-1999	99JUS-0142803
PR	09-JUL-1999	99JUS-0142920
PR	12-JUL-1999	99JUS-0142977
PR	13-JUL-1999	99JUS-0143542
PR	14-JUL-1999	99JUS-0143624
PR	15-JUL-1999	99JUS-0144005
PR	16-JUL-1999	99JUS-0144085
PR	16-JUL-1999	99JUS-0144086
PR	19-JUL-1999	99JUS-0144325
PR	19-JUL-1999	99JUS-0144331
PR	19-JUL-1999	99JUS-0144332
PR	19-JUL-1999	99JUS-0144333
PR	19-JUL-1999	99JUS-0144334
PR	20-JUL-1999	99JUS-0144335
PR	20-JUL-1999	99JUS-0144632
PR	20-JUL-1999	99JUS-0144884
PR	21-JUL-1999	99JUS-0144814
PR	21-JUL-1999	99JUS-0145086
PR	22-JUL-1999	99JUS-0145088
PR	22-JUL-1999	99JUS-0145085
PR	22-JUL-1999	99JUS-0145087
PR	22-JUL-1999	99JUS-0145089
PR	23-JUL-1999	99JUS-0145192
PR	23-JUL-1999	99JUS-0145145
PR	23-JUL-1999	99JUS-0145218
PR	26-JUL-1999	99JUS-0145224
PR	26-JUL-1999	99JUS-0145276
PR	27-JUL-1999	99JUS-0145918
PR	27-JUL-1999	99JUS-0145918
PR	27-JUL-1999	99JUS-0145919
PR	28-JUL-1999	99JUS-0145951
PR	02-AUG-1999	99JUS-0146386
PR	02-AUG-1999	99JUS-0146388
PR	03-AUG-1999	99JUS-0146389
PR	03-AUG-1999	99JUS-0147038
PR	04-AUG-1999	99JUS-0147204
PR	04-AUG-1999	99JUS-0147302
PR	05-AUG-1999	99JUS-0147192
PR	05-AUG-1999	99JUS-0147260
PR	06-AUG-1999	99JUS-0147303
PR	06-AUG-1999	99JUS-0147303
PR	06-AUG-1999	99JUS-0147416
PR	09-AUG-1999	99JUS-0147493
PR	09-AUG-1999	99JUS-0147935
PR	10-AUG-1999	99JUS-0148171
PR	11-AUG-1999	99JUS-0148319

```
QY 69 VTNEDGVKNVADNTVSTYGLKDIMFNSAGISDPNRPRIIDNEKADFERVLSVNVTVGVFLC 128
  ||| : |||| || : ||||| : ||||| ||| : ||| : ||| : |||
Db 75 VTKESDIANAVDFAVSLHTKLDIMYNNAGIPCKTPPSIVDLNLNVDKVINTRVGVMAG 134
  ||| : |||| || : ||||| : ||||| ||| : ||| : ||| : |||
QY 129 MKHAARVMIPARSGNIISTASLSTMTGGSSHAYCGSKHAVLALTRNLAVELGQFGIRYN 188
  : ||||| ||||| ||||| : ||| : ||| : ||| : ||| : ||| : |||
Db 135 IKAARVMIPRNSGSIICAGSVTGMGGLAQHTYSVSKSAVIGIVRSTASELCKHRIRYN 194
  : ||||| ||||| ||||| : ||| : ||| : ||| : ||| : ||| : |||
QY 189 CLSPFGLPTA-----LGKKFSGIKNEEFENVINFAGNLKGPKFNVEDYANAALYLASDE 243
  ||||| : ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 195 CISPPFAITTSFVMDENMRQIYPGV-DDSRLIQIVQSTGVNLNGEVCPTDVANAAYLASDD 253
  ||||| : ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 244 AKYVSGHNLFIDGGFSVCNSVIKVOYFP 271
  : ||||| ||||| : ||||| : : : : : : : : : : : : : : : : : : : : :
Db 254 SKYVNGHNLVVDGGF-----TTVKTLDFP 277
  : ||||| ||||| : ||||| : : : : : : : : : : : : : : : : : : : : :

RESULT 14
AAW06488
ID ABB92823 standard; Protein: 283 AA.
XX
AC ABB92823;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 2034.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
PA (FARB ) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
WPI; 2002-269010/31.
XX
DR
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
PS Claim 5; SEQ ID NO 2034; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
SQ Sequence 283 AA;

Query Match 42.9%; Score 597.5; DB 23; Length 283;
Best Local Similarity 47.0%; Pred. No. 6.4e-52;
Matches 126; Conservative 45; Mismatches 86; Indels 11; Gaps 4;

QY 9 RLLEKVALITGGASGIGETTAKLFQSHQCAKVAIADVQDELGHVSVEAIGTSNHYIICD 68
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 16 RKLEKVALITGGASGIGATKATGKFIHQHAKVLIADIQPIGRETEQELGPS-CAFFPCD 74
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 69 VTNEDGVKNVADNTVSTYGLKDIMFNSAGISDPNRPRIIDNEKADFERVLSVNVTVGVFLC 128
```

```
Db 75 VTKESDIANAVDFAVSLHTKLDIMYNNAGIPCKTPPSIVDLNLNVDKVINTRVGVMAG 134
  ||| : |||| || : ||||| : ||||| ||| : ||| : ||| : |||
QY 129 MKHAARVMIPARSGNIISTASLSTMTGGSSHAYCGSKHAVLALTRNLAVELGQFGIRYN 188
  : ||||| ||||| ||||| : ||| : ||| : ||| : ||| : ||| : |||
Db 135 IKAARVMIPRNSGSIICAGSVTGMGGLAQHTYSVSKSAVIGIVRSTASELCKHRIRYN 194
  : ||||| ||||| ||||| : ||| : ||| : ||| : ||| : ||| : |||
QY 189 CLSPFGLPTA-----LGKKFSGIKNEEFENVINFAGNLKGPKFNVEDYANAALYLASDE 243
  ||||| : ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 195 CISPPFAITTSFVMDENMRQIYPGV-DDSRLIQIVQSTGVNLNGEVCPTDVANAAYLASDD 253
  ||||| : ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 244 AKYVSGHNLFIDGGFSVCNSVIKVOYFP 271
  : ||||| ||||| : ||||| : : : : : : : : : : : : : : : : : : : : :
Db 254 SKYVNGHNLVVDGGF-----TTVKTLDFP 277
  : ||||| ||||| : ||||| : : : : : : : : : : : : : : : : : : : : :

RESULT 15
AAW06488
ID AAW06488 standard; Protein: 336 AA.
XX
AC AAW06488;
XX
DT 10-FEB-1997 (first entry)
XX
DE Maize Ts2 sequence.
XX
KW Tasselseed 2; Ts2; maize; H2 clone; rice; floral tissue; male; female;
KW bacterial hydroxysteroid dehydrogenase; sexual development; flower.
XX
OS Zea mays.
XX
PN WO9505732-A1.
XX
PD 02-MAR-1995.
XX
PF 23-AUG-1994; 94WO-US09513.
XX
PR 23-AUG-1993; 93US-0110690.
PA (UYVA ) UNIV YALE.
XX
PI Dellaporta SL;
XX
WPI; 1995-106591/14.
N-PSDB; AAT45060.
XX
Tasselseed 2 gene and promoter and expression systems for
transforming plants - for controlling sexual development to produce
either predominantly male or female flowers, to simplify
cross-breeding and production of hybrid seeds.
XX
Claim 2; Fig 6; 44pp; English.
XX
CC This sequence represents the tasselseed 2 (Ts2) protein derived from
CC maize. This sequence is encoded by the H2 clone. The Ts2 protein in
CC maize is very similar to Ts2 protein found in rice, and both these
CC sequences show considerable homology to bacterial hydroxysteroid
CC dehydrogenases. Ts2 has a mol. wt. of 35 kb and a predicted pI of
CC 6.7. The Ts2 coding sequence and its promoter, may be used to control
CC sexual development of floral tissue and to provide exclusively male
CC or female flowers.
XX
SQ Sequence 336 AA;

Query Match 42.2%; Score 588.5; DB 16; Length 336;
Best Local Similarity 42.2%; Pred. No. 6.8e-51;
Matches 122; Conservative 50; Mismatches 88; Indels 29; Gaps 4;

QY 5 TAFARLEKVALITGGASGIGETTAKLFQSHQCAKVAIADVQDELGHVSVEAIGTSNSTY 64
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47 TPMPEKLDGKVAIVTGGARGIGEIVRLFAKRGARVVIADIDDAAGEALASALGPQVS-F 105
  : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 65 IHCDVTNEDGVKNVADNTVSTY-GKLDIMFNSAGI---SDPNRPRIIDNEKADFERVLSV 120
```

Db 106 VRCDSVSEDDVRRRAVDWALSRHGRGLDVYCNNAAGVLGRQTRAARSILSFDAAEFDRVLRV 165  
Qy 121 NVTGVFLCMKHAARVMIIPARSGNIIISTASLSTMGSGSSHAYCGSKHAVLAL/TRLAVEL 180  
Db 166 NALGAALGMKHAARAMAPRAGSIVSVASVAVLGGLGPHAYTASKHAIVGLTKNAACEL 225  
Qy 181 GQFGIRVNCLSPECLPTA-----LGKKFSGIKNEEEFENV 216  
Db 226 RAHGRVNCVSPFGVATPMLINAWROGHDDATADADRLDLDLVTPSDQEVKMEEVV 285  
Qy 217 NFAGNLKGPKNVEDVANAALYLASDEAKYVSGHNLFTIDGGFSVCNSVI 265  
Db 286 RGLATLKGPTLRPRDIAEAVLFIADEARYISGHNLVVDGGVTTSRNLI 334

Search completed: February 14, 2003, 09:19:36  
Job time : 38 secs



GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2003, 09:20:43 : Search time 13 Seconds  
(without alignments)  
536.526 Million cell updates/sec

Title: us-09-673-918a-2

Perfect score: 1393

Sequence: 1 MQLRTAFARLEKGVKALITG.....IDGFGFVNSVIKVFQYFDS 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/FCI\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/FCIUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1233	88.5	276	9	US-09-944-160-24
2	432	31.0	253	10	US-09-815-242-11842
3	369	26.5	256	10	US-09-954-314-14
4	349.5	25.1	262	10	US-09-802-853-6
5	343.5	24.7	261	9	US-09-773-748-1
6	341.5	24.5	261	10	US-09-940-037A-29
7	332	23.8	263	10	US-09-815-242-14092
8	331	23.8	271	10	US-09-815-242-13387
9	331	23.8	272	10	US-09-815-242-13679
10	328.5	23.6	261	10	US-09-815-242-10846
11	328.5	23.6	272	10	US-09-815-242-12688
12	326	23.4	245	10	US-09-796-089-9
13	325	23.3	261	9	US-09-971-536-74
14	321.5	23.1	251	10	US-09-815-242-5689
15	321.5	23.1	270	9	US-09-978-295A-468
16	321.5	23.1	270	9	US-09-978-697-468
17	321.5	23.1	270	9	US-09-978-192A-468
18	321.5	23.1	270	9	US-09-999-832A-468
19	321.5	23.1	270	9	US-09-978-189-468

20	319.5	22.9	254	9	US-09-978-758-2	Sequence 2, Appli
21	319.5	22.9	258	9	US-10-156-055A-1	Sequence 1, Appli
22	318.5	22.9	246	10	US-09-739-254-143	Sequence 143, App
23	318.5	22.9	246	10	US-09-904-615-143	Sequence 143, App
24	318.5	22.9	253	10	US-09-931-186-15	Sequence 15, Appl
25	316.5	22.7	244	10	US-09-815-242-10126	Sequence 10126, A
26	310	22.3	285	10	US-09-815-242-10241	Sequence 10241, A
27	303	21.8	255	10	US-09-931-186-14	Sequence 14, Appl
28	301.5	21.6	246	10	US-09-815-242-5461	Sequence 5461, Ap
29	301.5	21.6	246	10	US-09-815-242-12123	Sequence 12123, A
30	301.5	21.6	246	10	US-09-815-242-12803	Sequence 12803, A
31	301.5	21.6	246	10	US-09-815-242-13100	Sequence 13100, A
32	296.5	21.3	262	10	US-09-815-242-10597	Sequence 10597, A
33	294.5	21.1	286	10	US-09-815-242-11860	Sequence 11860, A
34	288.5	20.7	274	9	US-09-976-059-17	Sequence 17, Appl
35	288	20.7	254	10	US-09-741-669-411	Sequence 411, App
36	287.5	20.6	242	10	US-09-815-242-10974	Sequence 10974, A
37	287	20.6	247	9	US-09-479-040-9	Sequence 9, Appli
38	283.5	20.4	245	10	US-09-815-242-10921	Sequence 10921, A
39	281	20.2	203	10	US-09-823-901-7	Sequence 7, Appli
40	277	19.9	258	9	US-09-738-626-6436	Sequence 6436, Ap
41	275	19.7	221	10	US-09-764-853-553	Sequence 553, App
42	274.5	19.7	206	9	US-09-997-816-4	Sequence 4, Appli
43	274.5	19.7	266	9	US-09-981-353-40	Sequence 40, Appl
44	274.5	19.7	266	10	US-09-931-186-18	Sequence 18, Appl
45	274	19.7	251	10	US-09-922-501-16	Sequence 16, Appl

## ALIGNMENTS

### RESULT 1

US-09-944-160-24  
Sequence 24, Application US/09944160  
Patent No. US20020174452A1  
GENERAL INFORMATION:  
APPLICANT: Lewis, No. US20020174452A1man  
APPLICANT: Davin, Laurence  
APPLICANT: .. Huang, Ning  
TITLE OF INVENTION: Monocot Seeds with Increased Lignan  
TITLE OF INVENTION: Content  
FILE REFERENCE: WSURL17983  
CURRENT APPLICATION NUMBER: US/09/944,160  
CURRENT FILING DATE: 2001-08-30  
PRIOR APPLICATION NUMBER: US 60/230,632  
PRIOR FILING DATE: 2000-09-07  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 24  
LENGTH: 276  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: secoisolariciresinol dehydrogenase amino acid  
OTHER INFORMATION: sequence from plasmid pAPI249  
US-09-944-160-24

Query Match	88.5%	Score	1233	DB	9	Length	276
Best Local Similarity	88.2%	Pred. No.	3.8e+108	Mismatches	14	Indels	0
Matches	240	Conservative	14	Gaps	0		
QY	2	QLRTAFARLEKGVKALITGASGIGETTAKLFSQHKAKVAIADVDQDELGHVSVEAIGTSN	61				
Db	5	QLRTAFARLEKGVKALITGASGIGETTAKLFSQHKAKVAIADVDQDELGHVSVEAIGLSN	64				
QY	62	STYIHCDVTNEDGVKNVADNTVSTYTKLDMFNSNAGISDPNRPRIIDNEKADFERVLSVN	121				
Db	65	STYIHCDVTNEDGVKNVADNTVSTYTKLDMFNNAGISDPYKPRVIDNEKADFERVLSVN	124				
QY	122	VTCVFLCMKHAARVMTIPARSGNIISTASLSSTMGSSSHAYCGSKHAVIALTRNLAVELG	181				
Db	125	VTCVFLCMKHAARVMTIPARSGNIISTASLSSTMGSSSHAYCGSKHAVIGLTRNLAVELG	184				



```
Best Local Similarity 31.2%; Pred. No. 2.2e-25;
Matches 83; Conservative 55; Mismatches 107; Indels 21; Gaps 6;

QY 8 ARLECKVALITGGASGIGETAKLFSSQHGAKVAIADVODEL--GHSVVEAIGTSNSTY 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 SKRFNGKCVLTGAGNIGLALRLAEETGTAIALDMMREALEKAEASVKGVEARSY 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 IHCVDVTNEDGVKNADVNTYSTYKGLDIMFSNAGISDPNRPRIIDNEKADFERVLSVNV 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 V-CDVTSEEAVTGVDSVVRDFEKIDFLENNAGYOGAFAP-VQDYPSDDFARVLTINV 119
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 VFLCMKHAARVMPARSIGNIISTASLSSTMGSSHAYCGSKHAVLALTRNLAVELGQ 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 AFHLKAVSRQMITQNYGRIVNTASAGYKPPNMAAYGASKGAIITALTETALDLP 179
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 IRVNCCLSP-----FGLPTALGKKFGIKNEEFENFVINFAGNLKGPKE-NVED 231
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 IRVNAISPGYMGPFMWERQVLOAKVGQSYFTDPKVVAAQMI----GSPMRRYGD 236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 232 VANAALYLASDEAKYVSGHNLFDGG 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 IPCVVAFLLDGDSSEFTGVNLPAGG 262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
US-09-773-748-1
; Sequence 1, Application US/09773748
; Publication No. US20020187537A1
; GENERAL INFORMATION:
; APPLICANT: Wada, Masaru
; TITLE OF INVENTION: Levodione reductase gene and use thereof
; FILE REFERENCE: Levodione reductase
; CURRENT APPLICATION NUMBER: US/09/773,748
; CURRENT FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: EP00101665.8
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1
; TYPE: PRT
; ORGANISM: Corynebacterium aquaticum
; US-09-773-748-1

Query Match 24.7%; Score 343.5; DB 9; Length 267;
Best Local Similarity 33.3%; Pred. No. 8.4e-25;
Matches 88; Conservative 48; Mismatches 105; Indels 23; Gaps 5;

QY 10 RLECKVALITGGASGIGETAKLFSSQHGAKVAIADVODE-----LGHVVEAIGTSNSTY 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 10 RETDRVLLITGGSGGLGRATAVRLAAEGAKLSLDVDSSEGLEASKAAVLETAPDAEVL 69
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 66 HCDVTNEDGVKNADVNTYSTYKGLDIMFSNAGISDPNRPRIIDNEK---ADFERVLSV 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 70 VADVDEAQAQVAYVTATTFERRIDGFFNNAGIEGKNP-----TESFTAEEFDKVV 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 TGVLFCMKHAARVMPARSIGNIISTASLSSTMGSSHAYCGSKHAVLALTRNLAVELGQ 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 RGVLGLEKVLKIMRQSGSMVNTASVGGIRGIGNQSGYAAKHGCVGLTRNSAVEYGR 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 FGIRVNCLSPFGLPTALGKKF-----SGIKNEEFENFVINFAGNLKGPKEFNVEDVANA 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 YGIRINAIAPGAIWTPMVENSKQLDPENPKRAAEFIQV-----NPSKRYGEAPEIAAV 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 236 ALYLASDEAKYVSGHNLFDGGFS 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 VAPLLSDASVYNTATVVPIDGGQS 264
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
US-09-940-037A-29
; Sequence 29, Application US/09940037A
; Patent No. US20020127679A1
```

```
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14092
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-14092

Query Match      23.8%; Score 332; DB 10; Length 263;
Best Local Similarity 33.1%; Pred. No. 9.9e-24;
Matches 87; Conservative 49; Mismatches 113; Indels 14; Gaps 5;

QY 10 RLEKVALITGGASGIGETTAKLFSQHGAKVAIADVQDELGHVSVEAIGTSN-STYIHCD 68
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3 KLTGKTALITGASQIGEGIARVFARHGANILLDISDEIEKLADELGGRGHRTAVKAD 62

QY 69 VYNEDGVKNVNTVSTYTKLIDIMFSNAGISDPNRPRIIDNEKADFERVLSVNTVGVFLC 128
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 63 VRDFASVQAARAKETEGRIDILVNNAGVCRNGFNLDMSSEEDRDFH--IDINIKGVWNV 120

QY 129 MKHAARVMIPARSGNIISTASLSTMGSSSHAYCGSKHVALTRNLNLADELGVGFGRV 187
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 TRAVLPFEMIKRKGGRVMMSSVTGDMVADPGETAAYALSAAIYGLTKSLAVEYAQSGIRV 180

QY 188 NCLSPFGLPTALGKFKSGIKNEEFENVINPAG-----NLGPKFKNVEDVANAALYLAS 241
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 NAICPGYVTRPMAESIAKQSPNDPESVLTEMAKAIPLRLADPL----EVGELAAFLAS 236

QY 242 DEAKVYSGHNLFDGGSVCNSV 264
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 237 DESSYLTPQNVIDGGSTLPESV 259

RESULT 8
US-09-815-242-13387
; Sequence 13387, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a US/09/815,242
; CURRENT APPLICATION NUMBER: 2001-03-21
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14092
; LENGTH: 263
; TYPE: PRT
; ORGANISM: streptococcus pneumoniae
US-09-815-242-13679
```

```
; SEQ ID NO 13387
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13387

Query Match      23.8%; Score 331; DB 10; Length 271;
Best Local Similarity 33.6%; Pred. No. 1.3e-23;
Matches 85; Conservative 48; Mismatches 110; Indels 10; Gaps 6;

QY 11 LEKQVALITGGASGIGETTAKLFSQHGAKVAIADVQDELGHVSVEA---IGTSNSTYIHC 67
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 13 LKQKALITGASVGIQFAIAKSAEAGATIVENDINQDLVKNKGEIAREVIEAIGV-C 71

QY 68 DVYNEDGVKNVNTVSTYTKLIDIMFSNAGISDPNRPRIIDNEKADFERVLSVNTVGVFL 127
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 72 DVTDEGIDQAMVKQIQEVEGVVIDILVNNAGII--RRVPMCEMSAADPKVIDIDLNAPFI 129

QY 128 CMKHAARVMIPARSGNIISTASLSTMGSSSHAYCGSKHVALTRNLNLADELGVGFGRV 187
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 VSKAVIPSMIKKGGHCKIINICSMMSSELGREITVSAYAAKGGCKMLTRNIASEYGGANIQC 189

QY 188 NCLSP--FGLP--TALGKFKSGIKNEEFENVINPAGNLGPKFKNVEDVANAALYLASDEA 244
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 190 NGIGPGYIATPQTAPLRELQEDGSRHPDFQFI-TAKTPAARWGNPDLMGPAVFLASDAS 248

QY 245 KYVSGHNLFDGGS 257
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 249 NFVNGHILYVDGG 261

RESULT 9
US-09-815-242-13679
; Sequence 13679, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13679
; LENGTH: 272
; TYPE: PRT
; ORGANISM: streptococcus pneumoniae
US-09-815-242-13679

Query Match      23.8%; Score 331; DB 10; Length 272;
Best Local Similarity 33.6%; Pred. No. 1.3e-23;
```

```

Db      69  FVGDVSSPSFVQATMKQIVKNDNGOIDLINTAGTILDDYRPS - LETSEALMDQIILATNLKS 127
QY      125  VFLCMKHAARVMIPARSGNIISTASLSSTMGSSSHAYCGSKHAVLALTRNLRAVELGQFG 184
      III : : : : : III: : : : : III: : : : :
Db      128  VFLVTNALLPYFLOOKKGVINMASIAGLVAGGGAAYTASKHAIICYTKOLSVDYAKLG 187
QY      185  IRVNCLSPPGLPTALGKKFGSKIKNEEFENVINPAGNLK-----GPKFNVEDVA 233
      III: : : : : I : : : : : I : : : : :
Db      188  IRANAIAPGAQTPTM-----NAADPAGEGEMATWVASETPAGRWAPQOEVA 233
QY      234  NAALYLASDEAKYVSGHNLFIDGGFSV 260
      : : : : : I : : : : :
Db      234  KLSLFASDDADYIHGVTMTIDGGWTM 260

RESULT 11
US-09-815-242-12688
; Sequence 12688, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; ; ; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12688
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12688

Query Match 23.6%; Score 328.5; DB 10; Length 272;
Best Local Similarity 30.1%; Pred. No. 2.2e-23;
Matches 78; Conservative 63; Mismatches 99; Indels 19; Gaps 7;

QY      9  RRLGKVALITGGASGTGTTAKLFSHQKAKVATADVODELGHV--VEAIGTNSNVIH 66
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2  KRLENKVAVVTGATGTGCGASATLAQEGAYVLAVDAIEAVSETVVDKIKSNGDKAKAVV- 60
QY      67  CDVTNEDGVKNAVDTVSTYCKLDIMESNAGISDPNRIIDNEKADFERVLSYNTGVF 126
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      61  VDIASEQIDINFASEIREQGHVDVLFENNAGV-DNAAGRHEYPFDVYDKIMNVDMRGTF 119
QY      127  LCMKHAARVMIP-----ARSGNIISTASLSSTMGSSSHAYCGSKHAVLALTRNLRAVELCOF 183
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      120  LMYK-----MMLPLMWTGKGSIVNTSSFGQADLYRSGYNAKCAVINFTKSIAYEGRD 175
QY      184  GIRVNCLSPPGLPTALGKKFGSKIKNEEFENVINPAGNLK-----GPKFNVEDVANAALY 238
      III : : : : : III: : : : : III: : : : : III: : : : :

```

Db 176 GIRANALAPGTIETPLVDKLTCTSEKKGK---AFRENQKWMTPILGRIGKPEEVGKLVVF 232  
QY 239 LASDEAKYVSGHNLFTDGG 257  
Db 233 LASDESSFTIGETIRIDGG 251

RESULT 12  
US-09-796-089-9  
; Sequence 9, Application US/09796089  
; Patent No. US2002010946A1  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; TITLE OF INVENTION: 21612, 21615, 21620, 21676, 33756, No. US20020010946A1el  
; FILE REFERENCE: 35800/208222  
; CURRENT APPLICATION NUMBER: US/09/796,089  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR FILING DATE: 2000-12-15  
; PRIOR FILING DATE: 1999-12-15  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 245  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-796-089-9

Query Match 23.4%; Score 326; DB 10; Length 245;  
Best Local Similarity 32.7%; Pred. No. 3 3e-23;  
Matches 83; Conservative 59; Mismatches 98; Indels 14; Gaps 7;  
QY 10 RLEGKVALITGASGIGETTAKLFSOHGAKVATADYQDELGHSVZAGTSNSTYIHCDY 69  
Db 3 RLDGKVILITAAAGIGQAALAFAREGAKVIATDINESKLQELKYPQIQTRV---LDV 59  
QY 70 TNEDEGVKNAVNTVSTYKGLDIMSAGISDPNRPRIIDNEKADFERVLSVNTGVFLCM 129  
Db 60 TK---KKQIDQFANEVERLDVLENVAGF--VHGVILDCEDKDWDFSNLNRSMYLM 113  
QY 130 KHAARVMIPARSGNIISTPASLSSTMTGGSSH-AVCGSKHAVLALTRNLAVELGQFGIRVN 188  
Db 114 KAPLPKLAQSGNLINMSSVASSVKGVNRCVYSTTKAAVIGLKSVAADPIQOGIRCN 173  
QY 189 CLSPFLPT-ALGKFKSGIKNEEFENVFNAGNLGPKF-NVEDVANAALYLASDEAKY 246  
Db 174 CVCPGTVDTPSLQERIQARGNPEARN--DFLRKQKTGRFAEELIAMLCLVILASDESAY 231  
QY 247 VSGHNLFTDGGFSV 260  
Db 232 VTGNPVIIDGGWSL 245

RESULT 13  
US-09-971-536-74  
; Sequence 74, Application US/09971536  
; Patent No. US20020159976A1  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Havukkala, Ilkka  
; APPLICANT: Bloksberg, Leonard  
; APPLICANT: Lubbers, Mark  
; APPLICANT: Dekker, James  
; APPLICANT: Christenson, Anna  
; APPLICANT: Holland, Ross  
; APPLICANT: O'Toole, Paul  
; APPLICANT: Reid, Julian  
; APPLICANT: Coolbear, Timothy  
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods  
; FILE REFERENCE: 1043c2

; CURRENT APPLICATION NUMBER: US/09/971,536  
; CURRENT FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238  
; PRIOR FILING DATE: 2000-08-08  
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160  
; PRIOR FILING DATE: 2001-08-08  
; NUMBER OF SEQ ID NOS: 83  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 74  
; LENGTH: 261  
; TYPE: PRT  
; ORGANISM: Lactobacillus rhamnosus  
US-09-971-536-74

Query Match 23.3%; Score 325; DB 9; Length 261;  
Best Local Similarity 33.8%; Pred. No. 4.4e-23;  
Matches 89; Conservative 51; Mismatches 99; Indels 24; Gaps 9;  
QY 9 RRLEGKVALITGASGIGETTAKLFSOHGAKVATADYQDELG-----HSVVEAIGTSNST 63  
Db 3 RDLNGKVAVVTGSGKIGAGIAEREGQEHMAVINYLGDHEGARKTADTVIKNGGOAVS- 61  
QY 64 YIHCDVTNEDGVKNAVNTVSTYKGLDIMSAG--ISDPNRPRIIDNEKADFERVLSVN 121  
Db 62 -IHADYSTAGIASLVKTAESFGRDLVYVNNAGMEIKAPTHEVSLD----DNKVIAIN 116  
QY 122 VTGVFLCMKHAARVMIP-ARSGNIISTASLSSTMTGGSSSHAYCGSKHAVLALTRNLAVEL 180  
Db 117 OTGVFLGARAALNYFLDHHQPGNIINISSVHEQIPWPTFASAAKGSVKLFTETIAMEY 176  
QY 181 GQFGRVNCISLFPGLPTAL-GKFKSGIKNEEFENVFN--AGNLGPKFNVEDVANAAL 237  
Db 177 ANGRVRNVAIGSGAETPINAEKFA--DKAQYDQTVAMIPQGRLGKP-----EDVAAGAA 229  
QY 238 YLASDEAKYVSGHNLFTDGGFSV 260  
Db 230 WLASTEYVVTGTTFLFDGGMTL 252

RESULT 14  
US-09-815-242-5689  
; Sequence 5689, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011a  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16

```
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5689
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5689

Query Match          23.1%; Score 321.5; DB 10; Length 251;
Best Local Similarity 30.0%; Pred. No. 8.9e-23;
Matches 77; Conservative 62; Mismatches 99; Indels 19; Gaps 7;

QY 11 LEKVALITGGASGIGETAKLPSQHGAKVATADVDELGHVS--VEATGTSNSTVIHCD 68
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 1 LENKVAVVTGASTGICQASAIQAQAGAYVLAVDIAEAVSETVDKIKSNGDKAKAYV-VD 59
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 69 VTNEDGVKNAVDTVTYKGLDIMFSNAGISDPNRRPIIDNEKADFVRLSVNVTGVFLC 128
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 60 IASEQQIDNPFASIEIQGFQGHVDVLFNAGV-DNAAGRIHEYPDVKIMNVDMRGTEFLM 118
||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 129 MKHAARVMIP---ARSGNIISTASLSSTMGSGSHAYCGSKHVALTRNLAVELGQFGI 185
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 119 TK---MMLPLMTKGSIVNTSFSQADLYRSYGYNAAKGAVINFTKSIAYEYGRDGI 174
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 186 RVNCLSPFGLPTALGKKFGSGIKNEEFENVINPAGNLK-----GPKFNVEDVANAALYLA 240
||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 175 RANAIAPGTIETPLVDKLTSEDKSGK---APRENQKWTPLGLRGKPEVGKLVVFLA 231
||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 241 SDEAKYVSGHNLFIDGG 257
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 232 SDESPFITGETIRIDGG 248

RESULT 15
US-09-978-295A-468
; Sequence 468, Application US/09978295A
; Patent No. US20020156006A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kijavin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James.
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2630P1C11
; CURRENT APPLICATION NUMBER: US/09/978,295A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/062250
```

```
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064249
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
```

1	PRIOR APPLICATION NUMBER: 60/081955
2	PRIOR FILING DATE: 1998-04-15
3	PRIOR APPLICATION NUMBER: 60/081917
4	PRIOR FILING DATE: 1998-04-15
5	PRIOR APPLICATION NUMBER: 60/081819
6	PRIOR FILING DATE: 1998-04-15
7	PRIOR APPLICATION NUMBER: 60/081952
8	PRIOR FILING DATE: 1998-04-15
9	PRIOR APPLICATION NUMBER: 60/081838
10	PRIOR FILING DATE: 1998-04-15
11	PRIOR APPLICATION NUMBER: 60/082568
12	PRIOR FILING DATE: 1998-04-21
13	PRIOR APPLICATION NUMBER: 60/082569
14	PRIOR FILING DATE: 1998-04-21
15	PRIOR APPLICATION NUMBER: 60/082704
16	PRIOR FILING DATE: 1998-04-22
17	PRIOR APPLICATION NUMBER: 60/082804
18	PRIOR FILING DATE: 1998-04-22
19	PRIOR APPLICATION NUMBER: 60/082700
20	PRIOR FILING DATE: 1998-04-22
21	PRIOR APPLICATION NUMBER: 60/082797
22	PRIOR FILING DATE: 1998-04-22
23	PRIOR APPLICATION NUMBER: 60/082796
24	PRIOR FILING DATE: 1998-04-23
25	PRIOR APPLICATION NUMBER: 60/083336
26	PRIOR FILING DATE: 1998-04-27
27	PRIOR APPLICATION NUMBER: 60/083322
28	PRIOR FILING DATE: 1998-04-28
29	PRIOR APPLICATION NUMBER: 60/083392
30	PRIOR FILING DATE: 1998-04-29
31	PRIOR APPLICATION NUMBER: 60/083495
32	PRIOR FILING DATE: 1998-04-29
33	PRIOR APPLICATION NUMBER: 60/083496
34	PRIOR FILING DATE: 1998-04-29
35	PRIOR APPLICATION NUMBER: 60/083499
36	PRIOR FILING DATE: 1998-04-29
37	PRIOR APPLICATION NUMBER: 60/083545
38	PRIOR FILING DATE: 1998-04-29
39	PRIOR APPLICATION NUMBER: 60/083554
40	PRIOR FILING DATE: 1998-04-29
41	PRIOR APPLICATION NUMBER: 60/083558
42	PRIOR FILING DATE: 1998-04-29
43	PRIOR APPLICATION NUMBER: 60/083559
44	PRIOR FILING DATE: 1998-04-29
45	PRIOR APPLICATION NUMBER: 60/083500
46	PRIOR FILING DATE: 1998-04-29
47	PRIOR APPLICATION NUMBER: 60/083742
48	PRIOR FILING DATE: 1998-04-30
49	PRIOR APPLICATION NUMBER: 60/084366
50	PRIOR FILING DATE: 1998-05-05
51	PRIOR APPLICATION NUMBER: 60/084414
52	PRIOR FILING DATE: 1998-05-06
53	PRIOR APPLICATION NUMBER: 60/084441
54	PRIOR FILING DATE: 1998-05-06
55	PRIOR APPLICATION NUMBER: 60/084637
56	PRIOR FILING DATE: 1998-05-07
57	PRIOR APPLICATION NUMBER: 60/084639
58	PRIOR FILING DATE: 1998-05-07
59	PRIOR APPLICATION NUMBER: 60/084640
60	PRIOR FILING DATE: 1998-05-07
61	PRIOR APPLICATION NUMBER: 60/084627
62	PRIOR FILING DATE: 1998-05-07
63	PRIOR APPLICATION NUMBER: 60/084643
64	PRIOR FILING DATE: 1998-05-07
65	PRIOR APPLICATION NUMBER: 60/085339
66	PRIOR FILING DATE: 1998-05-13
67	PRIOR APPLICATION NUMBER: 60/085338
68	PRIOR FILING DATE: 1998-05-13
69	PRIOR APPLICATION NUMBER: 60/085323

[illegible]

Search completed: February 14, 2003, 09:25:30  
Job time : 14 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 18:26:19 ; Search time 1468 seconds  
(without alignments)  
9035.493 Million cell updates/sec

Title: US-09-673-918a-1

Perfect score: 819

Sequence: 1 atgcagcttcgaactgcatt.....tgttccaatccagattct 819

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_nus:\*  
25: em\_gss\_Other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	2.6	457	13	BM359844 GA_Ea002
2	21	2.6	598	12	BG442276 GA_Ea001
3	21	2.6	615	12	BF269379 GA_ER000
4	21	2.6	780	17	AZ211531 SP_0155_B
5	21	2.6	911	12	BG444414 GA_Ea002
6	20	2.4	474	17	AQ790030 HS_3193_B

7	20	2.4	597	14	BM690984
8	20	2.4	601	17	AQ735999
9	20	2.4	638	12	BG166718
10	20	2.4	670	13	BF69871
11	20	2.4	786	12	BG714518
12	20	2.4	894	10	BE571531
13	19	2.3	426	10	AW856509
14	19	2.3	346	10	AV969254
15	19	2.3	473	17	BH737850
16	19	2.3	485	17	AQ185426
17	19	2.3	506	17	AQ514891
18	19	2.3	509	10	AW922556
19	19	2.3	515	12	BG157985
20	19	2.3	528	17	BH451022
21	19	2.3	537	13	BM318747
22	19	2.3	541	17	AZ622455
23	19	2.3	584	17	BH494130
24	19	2.3	587	17	BH555089
25	19	2.3	592	14	BP016339
26	19	2.3	593	12	BG048039
27	19	2.3	600	13	BM076927
28	19	2.3	600	17	AZ591300
29	19	2.3	635	17	AZ878058
30	19	2.3	643	17	AZ389227
31	19	2.3	645	17	BH569302
32	19	2.3	663	17	AQ315681
33	19	2.3	675	17	BH569310
34	19	2.3	700	14	BQ111969
35	19	2.3	702	14	BQ873587
36	19	2.3	704	14	BQ875148
37	19	2.3	705	17	BH603059
38	19	2.3	716	14	BQ972114
39	19	2.3	717	14	BQ971515
40	19	2.3	738	14	BQ870873
41	19	2.3	788	14	BQ506405
42	19	2.3	800	17	CNS034HP
43	19	2.3	809	17	AG160498
44	19	2.3	811	17	CNS0482C
45	19	2.3	821	17	CNS02LJ9

## ALIGNMENTS

RESULT 1  
BM359844  
LOCUS  
DEFINITION  
GA\_Ea0024E17r Gossypium arboreum 7-10 dpa fiber library Gossypium  
arborescens cDNA clone GA\_Ea0024E17r, mRNA sequence.  
ACCESSION  
BM359844  
VERSION  
BM359844.1  
KEYWORDS  
EST  
SOURCE  
Gossypium arboreum  
-ORGANISM  
Gossypium arboreum  
REFERENCE  
AUTHORS  
Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.  
TITLE  
An integrated analysis of the genetics, development, and evolution of the cotton fiber  
JOURNAL  
Unpublished (2000)  
COMMENT  
Contact: Wing RA  
Clemson University  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Total High Quality bases = 418  
Seq primer: TAATACGACTCACTAGG  
High quality sequence start: 5

FEATURES  
source High quality sequence stop: 451.  
Location/Qualifiers  
1. .457  
/organism="Gossypium arboreum"  
/strain="AKA"  
/cultivar="8400"  
/db\_xref="taxon:29729"  
/clone="GA\_Ea0024E17r"  
/clone\_lib="Gossypium arboreum 7-10 dpa fiber library"  
/tissue\_type="Fibers isolated from bolls harvested 7-10 dpa"  
/lab\_host="E. coli"  
/notes="Vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 116 a 131 c 122 g 87 t 1 others  
ORIGIN  
Query Match 2.6%; Score 21; DB 13; Length 457;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 CTACATCCACTGTGATGTTAC 209  
|||||  
Db 169 CTACATCCACTGTGATGTTAC 189

RESULT 2  
BG442276 598 bp mRNA linear EST 15-MAR-2001  
LOCUS GA\_Ea0016G23f Gossypium arboreum 7-10 dpa fiber library Gossypium  
DEFINITION arboreum cDNA clone GA\_Ea0016G23f, mRNA sequence.  
ACCESSION BG442276  
VERSION BG442276.1 GI:13351928  
KEYWORDS EST.  
SOURCE Gossypium arboreum.  
ORGANISM Gossypium arboreum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.  
1 (bases 1 to 598)  
AUTHORS Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry  
D., Wood, T.C., Leslie, A. and Wilkins, T.A.  
TITLE An integrated analysis of the genetics, development, and evolution  
of the cotton fiber  
JOURNAL Unpublished (2000)  
COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: TAATACGACTCTACTATAGG  
High quality sequence stop: 432.

FEATURES  
source 1. .598  
/organism="Gossypium arboreum"  
/strain="AKA"  
/cultivar="8400"  
/db\_xref="taxon:29729"  
/clone="GA\_Ea0016G23f"  
/clone\_lib="Gossypium arboreum 7-10 dpa fiber library"  
/tissue\_type="Fibers isolated from bolls harvested 7-10 dpa"  
/lab\_host="E. coli"  
/notes="Vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 161 a 119 c 200 g 117 t 1 others  
ORIGIN  
Query Match 2.6%; Score 21; DB 12; Length 598;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CATTGCTGATGCCAAGATGA 143

Db 316 CATTGCTGATGCCAAGATGA 336  
|||||

RESULT 3  
BF269379 615 bp mRNA linear EST 07-MAR-2001  
LOCUS GA\_EB0004I11f Gossypium arboreum 7-10 dpa fiber library Gossypium  
DEFINITION arboreum cDNA clone GA\_EB0004I11f, mRNA sequence.  
ACCESSION BF269379  
VERSION BF269379.1 GI:11200374  
KEYWORDS EST.  
SOURCE Gossypium arboreum.  
ORGANISM Gossypium arboreum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.  
1 (bases 1 to 615)  
AUTHORS Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry  
D., Wood, T.C., Leslie, A. and Wilkins, T.A.  
TITLE An integrated analysis of the genetics, development, and evolution  
of the cotton fiber  
JOURNAL Unpublished (2000)  
COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: TAATACGACTCTACTATAGG  
High quality sequence stop: 568.

FEATURES  
source 1. .615  
/organism="Gossypium arboreum"  
/strain="AKA"  
/cultivar="8400"  
/db\_xref="taxon:29729"  
/clone="GA\_EB0004I11f"  
/clone\_lib="Gossypium arboreum 7-10 dpa fiber library"  
/tissue\_type="Fibers isolated from bolls harvested 7-10 dpa"  
/lab\_host="E. coli"  
/notes="Vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 151 a 166 c 176 g 120 t 2 others  
ORIGIN  
Query Match 2.6%; Score 21; DB 12; Length 615;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 CTACATCCACTGTGATGTTAC 209  
|||||

Db 184 CTACATCCACTGTGATGTTAC 204

RESULT 4  
AZ211531 780 bp DNA linear GSS 31-AUG-2000  
LOCUS SF\_0155\_B1\_C09\_SP6E Strongylocentrotus purpuratus, purple sea  
DEFINITION urchin, sperm genomic BAC library Strongylocentrotus purpuratus  
genomic clone Plate=155 Col=17 Row=F, DNA sequence.  
ACCESSION AZ211531  
VERSION AZ211531.1 GI:8427406  
KEYWORDS GSS.  
SOURCE Strongylocentrotus purpuratus.

ORGANISM Strongylocentrotus purpuratus  
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
Echinoidea; Etechinoidea; Echinacea; Echinoida;  
Strongylocentrotidae; Strongylocentrotus.  
1 (bases 1 to 780)  
REFERENCE Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,  
Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray

TITLE  
JOURNAL  
MEDLINE  
COMMENT

/G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.  
A sea urchin genome project: Sequence scan, virtual map, and additional records  
Proc. Natl. Acad. Sci. U.S.A. 97 (17): 9514-9518 (2000)  
20402566  
Contact: Cameron, RA, Davidson, EH, Hood, L  
Division of Biology 156-29  
California Institute of Technology  
Pasadena California 91125, USA  
Tel: (626) 395-8421  
Fax: (626) 793-3047  
Email: acameron@caltech.edu  
Plate: 155 row: F column: 17  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 780.

FEATURES  
source  
Location/Qualifiers  
1. .780  
/organism="Strongylocentrotus purpuratus"  
/db\_xref="taxon:7668"  
/clone="plate:155 Col=17 Row=F"  
/clone\_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"  
/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli DH10B"

BASE COUNT 326 a 126 c 151 g 177 t

ORIGIN

Query Match 2.6%; Score 21; DB 17; Length 780;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 607 TTTTCAGGGATTAAATGAA 627  
|||||  
Db 97 TTTTCAGGGATTAAATGAA 117

RESULT 5  
BG444414 911 bp mRNA linear EST 15-MAR-2001  
LOCUS  
DEFINITION  
GA\_Ea0024E17f Gossypium arboreum 7-10 dpa fiber library Gossypium  
arboreum cDNA clone GA\_Ea0024E17f, mRNA sequence.

ACCESSION  
BG444414  
VERSION  
BG444414.1 GI:13354066  
KEYWORDS  
EST.  
SOURCE  
Gossypium arboreum.  
Gossypium arboreum.

ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.  
1 (bases 1 to 911)  
Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry  
D., Wood, T.C., Leslie, A. and Wilkins, T.A.  
An integrated analysis of the genetics, development, and evolution  
of the cotton fiber  
Unpublished (2000)  
Contact: Wing RA  
Clemson University  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: TAATACGACTCATATAGG  
High quality sequence stop: 650.

FEATURES  
source  
Location/Qualifiers  
1. .911  
/organism="Gossypium arboreum"  
/strain="AKA"  
/cultivar="8400"  
/db\_xref="taxon:29729"  
/clone="GA\_Ea0024E17f"

/clone\_lib="Gossypium arboreum 7-10 dpa fiber library"  
/tissue\_type="Fibers isolated from bolls harvested 7-10 dpa"  
/lab\_host="E. coli"  
/note="Vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 233 a 207 c 280 g 190 t

ORIGIN

Query Match 2.6%; Score 21; DB 12; Length 911;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 CTACATCCACTGTGATGTTAC 209  
|||||  
Db 170 CTACATCCACTGTGATGTTAC 190

RESULT 6  
AQ790030 474 bp DNA linear GSS 03-AUG-1999  
LOCUS  
DEFINITION  
HS\_3193\_B1\_E04\_MR CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate=3193 Col=7 Row=J, DNA sequence.

ACCESSION  
AQ790030  
VERSION  
AQ790030.1 GI:5697654  
KEYWORDS  
GSS.  
SOURCE  
human.  
Homo sapiens

ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 474)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallaceu.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: http://www.htsc.washington.edu  
Plate: 3193 row: J column: 7  
Seq primer: M13 Reverse  
Class: BAC ends  
High quality sequence stop: 474.

FEATURES  
source  
Location/Qualifiers  
1. 474  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="plate:3193 Col=7 Row=J"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 135 a 78 c 76 g 182 t

ORIGIN

Query Match 2.4%; Score 20; DB 17; Length 474;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 TCAACTATGGTGGTTC 473  
|||||  
Db 300 TCAACTATGGTGGTTC 319

RESULT 7  
BM690984



```
Plate: LLAM10227 row: d column: 10
High quality sequence stop: 636.
Location/Qualifiers
1. .638
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_89"
/tissue_type="hypernephroma, cell line"
/lab_host="DH10B (phage-resistant)"
/site="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 215 a 105 c 111 g 207 t
ORIGIN
Query Match 2.4%; Score 20; DB 12; Length 638;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 250 TCAACCTATGGAAGAACTGGA 269
|||||
Db 469 TCAACCTATGGAAGAACTGGA 488
|||||

RESULT 10
LOCUS BI669871 670 bp mRNA linear EST 12-SEP-2001
DEFINITION 603293467F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5312657 5',
mRNA sequence.
ACCESSION BI669871
VERSION BI669871.1 GI:15584104
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 670)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1192 row: f column: 18
High quality sequence stop: 670.
Location/Qualifiers
1. .670
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 213 a 123 c 126 g 208 t
ORIGIN
Query Match 2.4%; Score 20; DB 12; Length 638;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 250 TCAACCTATGGAAGAACTGGA 269
|||||
Db 431 TCAACCTATGGAAGAACTGGA 450
|||||

RESULT 12
LOCUS BE571531 894 bp mRNA linear EST 15-AUG-2000
DEFINITION 601333976F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3711403 5',
mRNA sequence.
```

```

ACCESSION   BE571531
VERSION     BE571531.1  GI:9815251
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus

REFERENCE   1 (bases 1 to 894)
AUTHORS    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL    National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT    Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM8935 row: 9 column: 20
High quality sequence stop: 651.
Location/Qualifiers
1..894
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3711403"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dn.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 278 a 127 c 190 g 299 t
ORIGIN
Query Match . 2.4%; Score 20; DB 10; Length 894;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 630 AGAATTTGAGATGTAATAA 649
|||||
Db 242 AGAATTTGAGATGTAATAA 261
|||||

RESULT 13
ACCESSION   AM856509
LOCUS       AM856509/c 346 bp mRNA linear EST 19-MAY-2000
DEFINITION PM4-CT0291-271099-001-d02 CT0291 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AM856509
VERSION     AM856509.1  GI:7952202
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 346)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics

ACCESSION   BE571531
VERSION     BE571531.1  GI:9815251
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus

REFERENCE   1 (bases 1 to 894)
AUTHORS    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL    National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT    Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM8935 row: 9 column: 20
High quality sequence stop: 651.
Location/Qualifiers
1..894
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3711403"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dn.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 278 a 127 c 190 g 299 t
ORIGIN
Query Match . 2.4%; Score 20; DB 10; Length 894;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 630 AGAATTTGAGATGTAATAA 649
|||||
Db 242 AGAATTTGAGATGTAATAA 261
|||||

RESULT 13
ACCESSION   AM856509
LOCUS       AM856509/c 346 bp mRNA linear EST 19-MAY-2000
DEFINITION PM4-CT0291-271099-001-d02 CT0291 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AM856509
VERSION     AM856509.1  GI:7952202
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 346)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-pM4-CT0291-271
099-001-4026t3-1999-10-27&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 25
High quality sequence stop: 137.
Location/Qualifiers
1..346
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0291"
/dev_stage="Adult"
/notes="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 65 a 82 c 96 g 103 t
ORIGIN
Query Match 2.3%; Score 19; DB 10; Length 346;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 GGAGCCAAAGTTGCCATTG 127
|||||
Db 113 GGAGCCAAAGTTGCCATTG 95
|||||

RESULT 14
ACCESSION   AV969254
LOCUS       AV969254 426 bp mRNA linear EST 14-MAR-2002
DEFINITION AV969254 Nori Satoh unpublished cDNA library, larva Ciona
intestinalis cDNA clone cily18907 5', mRNA sequence.
ACCESSION   AV969254
VERSION     AV969254.1  GI:19459018
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
ORGANISM    Ciona intestinalis.
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Ciona.
1 (bases 1 to 426)
Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.
Expressed genes in ciona intestinalis
Unpublished (2000)
Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoheascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1..426
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="cily18907"
/clone_lib="Nori Satoh unpublished cDNA library, larva"
/tissue_type="whole animal"
/dev_stage="larva"
/notes="Vector: pBluescript SK"

BASE COUNT 148 a 81 c 91 g 106 t
ORIGIN

```

Query Match 2.3%; Score 19; DB 10; Length 426;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 648 AACTTTTCGGGAAATTG 666  
|||||  
Db 367 AACTTTTCGGGAAATTG 385

RESULT 15  
BH737850  
LOCUS BH737850 473 bp DNA linear GSS 20-FEB-2002  
DEFINITION BOHZN64TR BO\_2\_3\_KB Brassica oleracea genomic clone BOHZN64, DNA  
sequence.  
ACCESSION BH737850  
VERSION BH737850.1 GI:18843245  
KEYWORDS GSS  
SOURCE Brassica oleracea.  
ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 473)  
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.  
TITLE Whole genome shotgun sequencing of Brassica oleracea  
JOURNAL Unpublished (2001)  
COMMENT Other\_GSSs: BOHZN64TF  
Contact: Chris Town

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: Sheared ends.

FEATURES  
source  
1..473  
/organism="Brassica oleracea"  
/strain="Tol1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHZN64"  
/clone\_lib="BO\_2\_3\_KB"  
/note="Vector: pHOS1; Site\_1: BstXI; 2-3 kb sheared  
genomic DNA inserted into pHOS1 using BstXI linkers"

BASE COUNT 147 a 76 c 97 g 153 t  
ORIGIN

Query Match 2.3%; Score 19; DB 17; Length 473;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 602 AGAAATTTTCAGGGATTAA 620  
|||||  
Db 192 AGAAATTTTCAGGGATTAA 210

Search completed: February 18, 2003, 19:41:21  
Job time : 1477 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 18:30:54 ; Search time 80 seconds  
(without alignments)  
3139.603 Million cell updates/sec

Title: US-09-673-918A-1

Perfect score: 819

Sequence: 1 atgcagcttcgaactgcatt.....tgtccaataatccagattctt 819

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A-COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B-COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A-COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B-COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS-COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	2.2	511	US-09-404-879A-64	Sequence 64, Appl
2	17	2.1	842	US-08-998-416-312	Sequence 312, App
3	17	2.1	3388	US-09-141-206-1	Sequence 1, Appli
4	17	2.1	3456	US-09-134-001C-405	Sequence 405, App
5	17	2.1	20137	US-09-262-773-206	Sequence 206, App
6	17	2.1	20138	US-09-262-773-9	Sequence 9, Appli
7	17	2.1	23071	US-09-262-773-210	Sequence 210, App
8	16	2.0	426	US-08-470-179-116	Sequence 116, App
9	16	2.0	480	US-09-134-001C-845	Sequence 845, App
10	16	2.0	515	US-09-222-575-143	Sequence 143, App
11	16	2.0	615	US-09-328-111-197	Sequence 197, App
12	16	2.0	861	US-09-134-001C-1266	Sequence 1266, Ap
13	16	2.0	1019	US-08-225-473-2	Sequence 2, Appli
14	16	2.0	1423	US-08-948-616-4	Sequence 4, Appli
15	16	2.0	1423	US-09-193-510-4	Sequence 4, Appli
16	16	2.0	1423	US-09-368-402-4	Sequence 4, Appli
17	16	2.0	1425	US-08-883-515-1	GENERAL INFORMAT
18	16	2.0	1839	US-09-134-001C-1191	Sequence 1191, Ap
19	16	2.0	2236	US-08-484-993B-13	Sequence 13, Appl
20	16	2.0	2236	US-08-484-158B-13	Sequence 13, Appl
21	16	2.0	2236	US-08-484-596A-13	Sequence 13, Appl
22	16	2.0	2236	US-08-480-150A-13	Sequence 13, Appl
23	16	2.0	2236	US-08-458-731-13	Sequence 13, Appl
24	16	2.0	2236	US-08-149-223A-13	Sequence 13, Appl
25	16	2.0	2337	US-09-156-253-1	Sequence 1, Appli
26	16	2.0	2558	US-08-999-733-3	Sequence 3, Appli
27	16	2.0	2571	US-07-718-575-7	Sequence 7, Appli

28	16	2.0	2971	1	US-08-481-206-7	Sequence 7, Appli
29	16	2.0	2971	2	US-08-486-269A-7	Sequence 7, Appli
30	16	2.0	3386	1	US-08-703-809-2	Sequence 2, Appli
31	16	2.0	3386	2	US-08-703-808-2	Sequence 2, Appli
32	16	2.0	3386	1	US-08-914-066-2	Sequence 2, Appli
33	16	2.0	3386	2	US-08-703-807-2	Sequence 2, Appli
34	16	2.0	3386	2	US-08-747-108A-2	Sequence 2, Appli
35	16	2.0	3386	3	US-09-211-631-2	Sequence 2, Appli
36	16	2.0	3386	3	US-09-265-628-2	Sequence 2, Appli
37	16	2.0	3386	4	US-09-001-141-2	Sequence 2, Appli
38	16	2.0	3386	4	US-09-532-803-2	Sequence 2, Appli
39	16	2.0	9103	1	US-08-466-033-182	Sequence 182, App
40	16	2.0	9103	1	US-08-444-733-182	Sequence 182, App
41	16	2.0	9103	2	US-08-464-134-182	Sequence 182, App
42	16	2.0	9103	2	US-08-461-361-182	Sequence 182, App
43	16	2.0	9103	2	US-08-485-910-182	Sequence 182, App
44	16	2.0	9103	5	PCT-US95-06286-156	Sequence 156, App
45	16	2.0	9757	1	US-08-093-453B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-09-404-879A-64  
; Sequence 64, Application US/09404879A  
; Patent No. 6468546  
; GENERAL INFORMATION:  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: King, Gordon E.  
; APPLICANT: Algate, Paul A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.462C2  
; CURRENT APPLICATION NUMBER: US/09/404,879A  
; CURRENT FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 393  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 64  
; LENGTH: 511  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-404-879A-64

Query Match 2.2%; Score 18; DB 4; Length 511;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 56 CAGGAGGAGCCAGTGGAA 73  
|||||  
DB 389 CAGGAGGAGCCAGTGGAA 406  
  
RESULT 2  
US-08-998-416-312  
; Sequence 312, Application US/089998416  
; Patent No. 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippsen, Peter  
; APPLICANT: Pohlmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jurgen  
; APPLICANT: Knechtle, Philipp  
; APPLICANT: Rebschung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: No. 6239264artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: No. 6239264th Carolina

```
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8589
; INFORMATION FOR SEQ ID NO: 312:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 842 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1253UP
; US-08-998-416-312

Query Match 2.1%; Score 17; DB 4; Length 842;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 CGAAGTCGATTCGCAAG 26
Db 42 CGAAGTCGATTCGCAAG 58

RESULT 3
US-09-141-206-1
; Sequence 1, Application US/09141206
; Patent No. 6187559
; GENERAL INFORMATION:
; APPLICANT: Steed, Paul M.
; APPLICANT: Lasala, Daniel J.
; TITLE OF INVENTION: Amino Acid Sequence of Human PLD2A
; Patent No. 6187559
; FILE REFERENCE: 4-30148/PL/CGC1954/R
; CURRENT APPLICATION NUMBER: US/09/141,206
; CURRENT FILING DATE: 1998-08-27
; EARLIER APPLICATION NUMBER: 60/057,802
; EARLIER FILING DATE: 1997-08-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3388
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (162)..(2963)
; OTHER INFORMATION: Human PLD2
; US-09-141-206-1

Query Match 2.1%; Score 17; DB 4; Length 3388;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 ACTGGACATTATGTCA 280
Db 10849 GGTGTCCTCAATATCCAGA 10833

US-09-134-001C-405
; Sequence 405, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 405
; LENGTH: 3456
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-405

Query Match 2.1%; Score 17; DB 4; Length 3456;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 674 CAAAATTTAATGTTGAG 690
Db 3095 CAAAATTTAATGTTGAG 3111

RESULT 5
US-09-262-773-206/C
; Sequence 206, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; SUSCEPTIBILITY GENE CHD1
; FILE REFERENCE: Myriad 3
; CURRENT APPLICATION NUMBER: US/09/262,773
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 20137
; TYPE: DNA
; ORGANISM: human
; US-09-262-773-206

Query Match 2.1%; Score 17; DB 4; Length 20137;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 799 GGTGTCCTCAATATCCAGA 815
Db 10849 GGTGTCCTCAATATCCAGA 10833

RESULT 6
US-09-262-773-9/c
; Sequence 9, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
```

; APPLICANT: Hess, Mark A.  
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE  
; FILE REFERENCE: Myriad 3  
; CURRENT APPLICATION NUMBER: US/09/262,773  
; CURRENT FILING DATE: 1999-03-04  
; NUMBER OF SEQ ID NOS: 210  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 20138  
; TYPE: DNA  
; ORGANISM: human  
US-09-262-773-9

Query Match 2.1%; Score 17; DB 4; Length 20138;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 799 GTGTTCCATATCCAGA 815  
|||||  
Db 10850 GTGTTCCATATCCAGA 10834

RESULT 7  
US-09-262-773-210/c  
; Sequence 210, Application US/09262773  
; Patent No. 6225451  
; GENERAL INFORMATION:  
; APPLICANT: Ballinger, Dennis G.  
; APPLICANT: Ding, Wei  
; APPLICANT: Wagner, Susanne  
; APPLICANT: Hess, Mark A.  
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE  
; FILE REFERENCE: Myriad 3  
; CURRENT APPLICATION NUMBER: US/09/262,773  
; CURRENT FILING DATE: 1999-03-04  
; NUMBER OF SEQ ID NOS: 210  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 210  
; LENGTH: 23071  
; TYPE: DNA  
; ORGANISM: human  
US-09-262-773-210

Query Match 2.1%; Score 17; DB 4; Length 23071;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 799 GTGTTCCATATCCAGA 815  
|||||  
Db 13783 GTGTTCCATATCCAGA 13767

RESULT 8  
US-08-470-179-116/c  
; Sequence 116, Application US/08470179  
; Patent No. 5645994  
; GENERAL INFORMATION:  
; APPLICANT: Huang Ph.D, Wai Mun  
; TITLE OF INVENTION: Method and Compositions for  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Trask, Britt and Rossa  
; STREET: P.O. Box 2550  
; CITY: Salt Lake City  
; STATE: Utah  
; COUNTRY: USA  
; ZIP: 84110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,179  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sweigert Ph.D, Susan E.  
; REGISTRATION NUMBER: 36,289  
; REFERENCE/DOCKET NUMBER: 2601  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 801-532-1922  
; TELEFAX: 801-531-9168  
; INFORMATION FOR SEQ ID NO: 116:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 426 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Anacystis nidulans  
US-08-470-179-116

Query Match 2.0%; Score 16; DB 1; Length 426;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GAGCCAAAGTTGCCAT 125  
|||||  
Db 218 GAGCCAAAGTTGCCAT 203

RESULT 9  
US-09-134-001C-845  
; Sequence 845, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 845  
; LENGTH: 480  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-845

Query Match 2.0%; Score 16; DB 4; Length 480;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 793 ATCAAAAGTTGTTCAAT 808  
|||||  
Db 335 ATCAAAAGTTGTTCAAT 350

RESULT 10  
US-09-222-575-143/c  
; Sequence 143, Application US/09222575  
; Patent No. 6387697  
; GENERAL INFORMATION:  
; APPLICANT: Yuqiu, Jlang  
; APPLICANT: Dillon, Davin C.

```
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: Compositions for the Treatment and Diagnosis of Breast Cancer
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 210121.470
; CURRENT APPLICATION NUMBER: US/09/222,575
; CURRENT FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 143
; LENGTH: 515
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (5)
; OTHER INFORMATION: where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (286)
; OTHER INFORMATION: where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (498)
; OTHER INFORMATION: where n is a, c, g or t
; US-09-222-575-143

Query Match          2.0%; Score 16; DB 4; Length 515;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 623 ATGAAGAGAAATTTCG 638
Db 281 ATGAAGAGAAATTTCG 266
|||||

RESULT 11
US-09-328-111-197/c
; Sequence 197, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Cartoll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 197
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(615)
; OTHER INFORMATION: n = A,T,C or G
; US-09-328-111-197

Query Match          2.0%; Score 16; DB 4; Length 615;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 271 ATTATGTTCAAGCATG 286
Db 261 ATTATGTTCAAGCATG 246
|||||

RESULT 12
US-09-134-001C-1266
; Sequence 1266, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1266
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-1266

Query Match          2.0%; Score 16; DB 4; Length 861;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 ATTAATAATGAAGAAG 631
Db 676 ATTAATAATGAAGAAG 691
|||||

RESULT 13
US-08-225-473-2/c
; Sequence 2, Application US/08225473
; Patent No. 5610012
; GENERAL INFORMATION:
; APPLICANT: Luchansky, John B.
; APPLICANT: Chen, Jianchi
; TITLE OF INVENTION: DNA Probes Specific For Virulent
; TITLE OF INVENTION: Listeria Monocytogenes
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/225,473
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Muetting, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 800.16-US-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1019 base pairs
```

;  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
US-08-225-473-2

Query Match 2.0%; Score 16; DB 1; Length 1019;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AACTGCATTCGCAAGA 27  
|||||  
Db 157 AACTGCATTCGCAAGA 142

## RESULT 14

US-08-948-616-4  
; Sequence 4, Application US/08948616  
; Patent No. 5840539

; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Lal, Preeti  
; APPLICANT: Shah, Purvi  
; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS  
; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA

; COUNTRY: USA  
; ZIP: 94304

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/948,616  
; FILING DATE: Herewith

; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:  
; FILING DATE:

; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0409 US

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166  
; TELEX:

; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 1423 base pairs  
; TYPE: nucleic acid

; STRANDEDNESS: single  
; TOPOLOGY: linear

; IMMEDIATE SOURCE:  
; LIBRARY: LUNGTUT07

; CLONE: 2607662  
US-08-948-616-4

Query Match 2.0%; Score 16; DB 2; Length 1423;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 GAAACTGGACATTAT 275  
|||||  
Db 204 GAAACTGGACATTAT 219

## RESULT 15

US-09-193-510-4

; Sequence 4, Application US/09193510  
; Patent No. 5981226

; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Lal, Preeti  
; APPLICANT: Shah, Purvi  
; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS  
; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA

; COUNTRY: USA  
; ZIP: 94304

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/193,510  
; FILING DATE:

; CLASSIFICATION:  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/948,616  
; FILING DATE:

; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0409 US

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166  
; TELEX:

; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 1423 base pairs  
; TYPE: nucleic acid

; STRANDEDNESS: single  
; TOPOLOGY: linear

; IMMEDIATE SOURCE:  
; LIBRARY: LUNGTUT07

; CLONE: 2607662  
US-09-193-510-4

Query Match 2.0%; Score 16; DB 2; Length 1423;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 GAAACTGGACATTAT 275  
|||||

Db 204 GAAACTGGACATTAT 219

Search completed: February 18, 2003, 19:43:18  
Job time : 113 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 19:33:55 ; Search time 114 Seconds  
(without alignments)  
3659.088 Million cell updates/sec

Title: US-09-673-918A-1

Perfect score: 819

Sequence: 1 atgcagcttcgaactgcatt.....tgtccaataatccagattctt 819

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 424239 seqs, 254661826 residues

Word size : 0

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published\_Applications\_NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	11.1	4505	9 US-09-944-160-23	Sequence 23, Appl
2	21	2.6	21	9 US-09-944-160-33	Sequence 33, Appl
3	21	2.6	21	9 US-09-944-160-43	Sequence 43, Appl
4	19	2.3	386	10 US-09-960-352-14577	Sequence 14577, A
5	18	2.2	292	9 US-09-796-692-5749	Sequence 5749, Ap
6	18	2.2	480	10 US-09-867-701-1789	Sequence 1789, Ap
7	18	2.2	511	10 US-09-884-441-64	Sequence 64, Appl
8	18	2.2	2016	10 US-09-764-870-110	Sequence 237, App
9	18	2.2	2223	10 US-09-764-870-110	Sequence 110, App
10	18	2.2	10785	10 US-09-764-878-324	Sequence 324, App
11	18	2.2	13315	10 US-09-764-878-328	Sequence 328, App
12	18	2.2	20261	10 US-09-764-878-325	Sequence 325, App
13	18	2.2	35641	10 US-09-962-436-306	Sequence 306, App
14	18	2.2	35641	10 US-09-880-107-2225	Sequence 2225, App
15	17	2.1	240	10 US-09-923-876-3491	Sequence 3491, Ap
16	17	2.1	356	10 US-09-960-352-12402	Sequence 12402, A
17	17	2.1	425	10 US-09-880-107-2010	Sequence 2010, Ap
18	17	2.1	517	10 US-09-880-107-1511	Sequence 1511, Ap
19	17	2.1	554	10 US-09-864-761-11940	Sequence 11940, A

c 20	17	2.1	596	10 US-09-864-761-16202	Sequence 16202, A
c 21	17	2.1	708	10 US-09-770-149-223	Sequence 223, App
c 22	17	2.1	756	9 US-09-774-639-64	Sequence 64, Appl
c 23	17	2.1	1438	10 US-09-884-901-1	Sequence 1, Appl1
c 24	17	2.1	5313	10 US-09-801-368-297	Sequence 297, App
c 25	17	2.1	16423	9 US-09-989-442-160	Sequence 160, App
c 26	17	2.1	38059	10 US-09-880-107-2125	Sequence 2125, Ap
c 27	17	2.1	513509	9 US-09-754-853A-4	Sequence 4, Appl1
c 28	16	2.0	100	10 US-09-864-761-32694	Sequence 32694, A
c 29	16	2.0	213	10 US-09-878-574-5029	Sequence 5029, Ap
c 30	16	2.0	270	10 US-09-878-574-12477	Sequence 12477, A
c 31	16	2.0	271	10 US-09-878-574-5697	Sequence 5697, App
c 32	16	2.0	291	10 US-09-878-574-7983	Sequence 7983, Ap
c 33	16	2.0	300	10 US-09-783-590-1117	Sequence 1117, Ap
c 34	16	2.0	319	10 US-09-864-761-23571	Sequence 23571, A
c 35	16	2.0	328	10 US-09-969-708-572	Sequence 572, App
c 36	16	2.0	373	10 US-09-983-965-4591	Sequence 4591, Ap
c 37	16	2.0	379	10 US-09-925-239-124	Sequence 124, App
c 38	16	2.0	380	10 US-09-983-965-215	Sequence 215, App
c 39	16	2.0	401	9 US-10-025-380-113	Sequence 113, App
c 40	16	2.0	401	10 US-09-922-217-113	Sequence 113, App
c 41	16	2.0	401	10 US-09-833-263-113	Sequence 113, App
c 42	16	2.0	401	10 US-09-878-574-3924	Sequence 3924, Ap
c 43	16	2.0	411	10 US-09-878-574-4099	Sequence 4099, Ap
c 44	16	2.0	412	9 US-09-954-531-790	Sequence 790, App
c 45	16	2.0	412	10 US-09-954-456-893	Sequence 893, App

ALIGNMENTS

RESULT 1  
US-09-944-160-23  
; Sequence 23, Application US/09944160  
; Patent No. US20020174452A1  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, No. US20020174452A1man  
; APPLICANT: Davin, Laurence  
; APPLICANT: .. Huang, Ning  
; TITLE OF INVENTION: Monocot Seeds with Increased Lignan  
; TITLE OF INVENTION: Content  
; FILE REFERENCE: WSURL17983  
; CURRENT APPLICATION NUMBER: US/09/944,160  
; CURRENT FILING DATE: 2001-08-30  
; PRIOR APPLICATION NUMBER: US 60/230,632  
; PRIOR FILING DATE: 2000-09-07  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 23  
; LENGTH: 4505  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: plasmid pAPI249 encoding secoisolariciresinol  
; OTHER INFORMATION: dehydrogenase (dehydrogenase)  
US-09-944-160-23

Query Match 11.1%; Score 91; DB 9; Length 4505;  
Best Local Similarity 100.0%; Pred.No. 9.2e-40;  
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CACAGCAAACTCTTCTCCCAACATGGCCAAAGTTGCCATTGCTGCATGCCAAGATGA 143

Db 1045 CACAGCAAACTCTTCTCCCAACATGGCCAAAGTTGCCATTGCTGCATGCCAAGATGA 1104

QY 144 ATTAGTCTACTCAGTTGTTCGAGCCCATTTGGC 174

Db 1105 ATTAGTCTACTCAGTTGTTCGAGCCCATTTGGC 1135

RESULT 2

US-09-944-160-33/c

; Sequence 33, Application US/09944160

```
; Patent No. US20020174452A1
; GENERAL INFORMATION:
; APPLICANT: Lewis, No. US20020174452Alman
; APPLICANT: Davis, Laurence
; APPLICANT: Huang, Ning
; TITLE OF INVENTION: Monocot Seeds with Increased Lignan
; FILE REFERENCE: WSURL17983
; CURRENT APPLICATION NUMBER: US/09/944,160
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 60/230,632
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer
US-09-944-160-33

Query Match      2.6%; Score 21; DB 9; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 GGGTGGTGGTCTTCACATGC 482
DB 21 GGGTGGTGGTCTTCACATGC 1

RESULT 3
US-09-944-160-43/c
; Sequence 43, Application US/09944160
; Patent No. US20020174452A1
; GENERAL INFORMATION:
; APPLICANT: Lewis, No. US20020174452Alman
; APPLICANT: Davis, Laurence
; APPLICANT: .. Huang, Ning
; TITLE OF INVENTION: Monocot
; FILE REFERENCE: WSURL17983
; CURRENT APPLICATION NUMBER: US/09/944,160
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 60/230,632
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer
US-09-944-160-43

Query Match      2.6%; Score 21; DB 9; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 GGGTGGTGGTCTTCACATGC 482
DB 21 GGGTGGTGGTCTTCACATGC 1

RESULT 4
US-09-960-352-14577/c
; Sequence 14577, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; PRIOR FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14577
; LENGTH: 386
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 62-LIB34-012-Q1-E1-H10
US-09-960-352-14577

Query Match      2.3%; Score 19; DB 10; Length 386;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 793 ATCAAAGTGTTCACATATC 811
DB 46 ATCAAAGTGTTCACATATC 28

RESULT 5
US-09-796-692-5749/c
; Sequence 5749, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Jane
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5749
; LENGTH: 292
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-5749

Query Match      2.2%; Score 18; DB 9; Length 292;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 AAGCTAGAGGAAAGT 44
```



```
Db 42 AAGCTAGTAAAGGAAAAGT 25
|||||
RESULT 6
US-09-867-701-1789
; Sequence 1789, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1789
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(480)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-1789

Query Match 2.2%; Score 18; DB 10; Length 480;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGAGGAGCCAGTGGAAAT 75
|||||
Db 424 GGAGGAGCCAGTGGAAAT 441

RESULT 7
US-09-884-441-64
; Sequence 64, Application US/09884441
; Patent No. US20020119158A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C7
; CURRENT APPLICATION NUMBER: US/09/884,441
; CURRENT FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 511
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-884-441-64

Query Match 2.2%; Score 18; DB 10; Length 511;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 CAGGAGGAGCCAGTGGAA 73
|||||
Db 389 CAGGAGGAGCCAGTGGAA 406

RESULT 8
US-09-764-870-237/c
; Sequence 237, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
```

---

```
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 237
; LENGTH: 2016
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (37)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (48)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (93)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-870-237

Query Match 2.2%; Score 18; DB 10; Length 2016;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGAGGAGCCAGTGGAAAT 75
|||||
Db 1098 GGAGGAGCCAGTGGAAAT 1081

RESULT 9
US-09-764-870-110/c
; Sequence 110, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 110
; LENGTH: 2223
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-870-110

Query Match 2.2%; Score 18; DB 10; Length 2223;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGAGGAGCCAGTGGAAAT 75
|||||
Db 1249 GGAGGAGCCAGTGGAAAT 1232

RESULT 10
US-09-764-878-324/c
; Sequence 324, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 324
; LENGTH: 10785
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-324

Query Match      2.2%; Score 18; DB 10; Length 10785;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGAGGAGCCAGTGGAAATT 75
Db 9833 GGAGGAGCCAGTGGAAATT 9816
|||||

RESULT 11
US-09-764-878-328/c
; Sequence 328, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 328
; LENGTH: 13315
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-328

Query Match      2.2%; Score 18; DB 10; Length 13315;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGAGGAGCCAGTGGAAATT 75
Db 12363 GGAGGAGCCAGTGGAAATT 12346
|||||

RESULT 12
US-09-764-878-325/c
; Sequence 325, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 325
; LENGTH: 20261
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-325

Query Match      2.2%; Score 18; DB 10; Length 20261;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGAGGAGCCAGTGGAAATT 75
Db 19310 GGAGGAGCCAGTGGAAATT 19293
|||||

RESULT 13
US-09-962-436-306

; Sequence 306, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,082
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 306
; LENGTH: 35641
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-436-306

Query Match      2.2%; Score 18; DB 10; Length 35641;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 608 TTTCAGGAGTAAAAAATG 625
Db 19637 TTTCAGGAGTAAAAAATG 19654
|||||

RESULT 14
US-09-880-107-2225
; Sequence 2225, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherif, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2225
; LENGTH: 35641
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L76568
US-09-880-107-2225

Query Match      2.2%; Score 18; DB 10; Length 35641;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 608 TTTCAGGAGTAAAAAATG 625
Db 19637 TTTCAGGAGTAAAAAATG 19654
|||||

RESULT 15
US-09-923-876-3491/c
; Sequence 3491, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
```

APPLICANT: Sherman, Bradley K.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING  
FILE REFERENCE: PL-0012-1 CON  
CURRENT APPLICATION NUMBER: US/09/923,876  
CURRENT FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: 09/298,329  
PRIOR FILING DATE: 1999-04-21  
PRIOR APPLICATION NUMBER: 60/085,331  
PRIOR FILING DATE: 1998-05-05  
NUMBER OF SEQ ID NOS: 6332  
SOFTWARE: PERL Program  
SEQ ID NO 3491  
LENGTH: 240  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: misc.feature  
OTHER INFORMATION: Incyte ID No. US20020013958A1 700162378H1  
US-09-923-876-3491

Query Match 2.1%; Score 17; DB 10; Length 240;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 TTCAGCAATGCAGGAAT 293  
|||||  
Db 92 TTCAGCAATGCAGGAAT 76

Search completed: February 18, 2003, 20:29:04  
Job time : 160 secs

**THIS PAGE BLANK (USPTO)**

Result No.	Query Match	Score	Length	DB ID	Description
1	100.0	819	21	AA245730	Secoisolariciresin
2	816	99.6	816	AA245743	Secoisolariciresin
3	647	79.0	831	AA245731	Secoisolariciresin
4	95	11.6	819	AA245732	Secoisolariciresin
5	91	11.1	828	AA245734	Secoisolariciresin
6	91	11.1	4505	AA384332	Plasmid pAP1249 en
7	87	10.6	831	AA245733	Secoisolariciresin
8	29	3.5	33	AA245744	PCR primer DEHY130
9	23	2.8	21	AA245741	PCR primer DEHY130

RESULT 1	
AAZ45730	
ID AAZ45730 standard; cDNA; 819 BP.	
XX AC	
XX AAZ45730;	
XX DT	
XX 06-APR-2000 (first entry)	
XX DE	
XX DE Secoisolaricresinol dehydrogenase cDNA clone DERV133.	
XX KW Secoisolaricresinol dehydrogenase; lignan biosynthetic pathway;	
XX KW lignan; matairesinol; health-protecting lignan; phytoestrogen;	
XX KW enterolactone; enterodiol; neutraceutical; dietary supplement;	
XX KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.	
XX OS Forsythia intermedia.	
XX FT	
XX Key Location/Qualifiers	
CDS 1..819	
FT /tag= a	
FT /product= "secoisolaricresinol dehydrogenase"	
FT /note= "no termination codon given"	
XX	
PN W09955846-Al.	
XX	
XX PD	
XX 04-NOV-1999.	
PF 23-APR-1999; 99WO-US08975.	
XX	
PPR 24-APR-1998; 98US-0082977.	
XX	
PA (UNIW ) UNIV WASHINGTON STATE RES. FOUND.	

XX PI Xia Z, Costa MA, Davin LB, Lewis NG;  
 XX WPI: 2000-126356/11.  
 XX P-PSDB: AAY54412.  
 XX New nucleic acid molecule encoding an enzyme involved in lignan  
 PT biosynthetic pathway, useful for producing large amounts of lignans  
 PS Claim 9; Page 45-46; 66pp; English.  
 XX The present sequence encodes a secoisolariciresinol dehydrogenase  
 CC protein. The enzyme is involved in the lignan biosynthetic pathway.  
 CC The secoisolariciresinol dehydrogenase proteins have a molecular  
 CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require  
 CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase  
 CC nucleic acids are used for the recombinant expression of the enzymes.  
 CC It is also used to obtain expression or enhanced expression of  
 CC secoisolariciresinol dehydrogenase in plants or to alter lignan  
 CC biosynthesis. The enzyme is used for production of the pharmacologically  
 CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase  
 CC proteins and nucleic acids can be utilized to: elevate or otherwise  
 CC alter the levels of health-protecting lignans, including phytoestrogens  
 CC such as enterolactone and enterodiol, in plant species, including  
 CC vegetables, grains and fruits and to food items incorporating material  
 CC derived from such genetically altered plants; genetically alter plant  
 CC species to provide an abundant, natural supply of lignans useful for  
 CC a variety of purposes, for example as nutraceuticals and dietary  
 CC supplements; to genetically alter living organisms to produce an  
 CC abundant supply of optically pure lignans having desirable biological  
 CC properties, for example (-)-trachelogenin which possesses antiviral  
 CC properties, and (-)-podophyllotoxin.  
 XX  
 SQ Sequence 819 BP; 240 A; 166 C; 189 G; 224 T; 0 other;  
 Query Match 100.0%; Score 819; DB 21; Length 819;  
 Best Local Similarity 100.0%; Pred No. 0;  
 Matches 819; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTTCCAACTGATTCGCAAGAGCTAGAGAGAAAGTTGCCCTTATACAGGA 60  
 DB 1 ATGCAGCTTCCAACTGATTCGCAAGAGCTAGAGAGAAAGTTGCCCTTATACAGGA 60  
 QY 61 GGAGCCAGTGAATTGGAGAACACACAGCAAAACTCTTCCCAACATGGAGCCAAAGTT 120  
 DB 61 GGAGCCAGTGAATTGGAGAACACACAGCAAAACTCTTCCCAACATGGAGCCAAAGTT 120  
 QY 121 GCCATTGCTGATGTCCCAAGATGAATTAGGTCACTCACTAGTTGTCGAGGCCATTGGCACTTCC 180  
 DB 121 GCCATTGCTGATGTCCCAAGATGAATTAGGTCACTCACTAGTTGTCGAGGCCATTGGCACTTCC 180  
 QY 181 AATTCCACCTACATCCACTGTGATGTACTTAATGAAGACGGTGTAAATAATGCCGTGGAC 240  
 DB 181 AATTCCACCTACATCCACTGTGATGTACTTAATGAAGACGGTGTAAATAATGCCGTGGAC 240  
 QY 241 AACACAGTTTCAACCTATGGAACATGAGACAGAAAGAGACACTTTGAACGGTCTCAGTGTA 300  
 DB 241 AACACAGTTTCAACCTATGGAACATGAGACAGAAAGAGACACTTTGAACGGTCTCAGTGTA 300  
 QY 301 CCCACAGGCCCCCATCATAGACAGAAAGAGACACTTTGAACGGTCTCAGTGTA 360  
 DB 301 CCCACAGGCCCCCATCATAGACAGAAAGAGACACTTTGAACGGTCTCAGTGTA 360  
 QY 361 AATGTACCGGAGTTTCCCTATGATGATGAACGACGACGACGAGTGTATGATTCACGACGC 420  
 DB 361 AATGTACCGGAGTTTCCCTATGATGATGAACGACGACGACGAGTGTATGATTCACGACGC 420  
 QY 421 AGTGGCAACATAATTTCCACTGCTAGTATTAAGCTCAACTATGGGTGGTGGTCTTCACAT 480  
 DB 421 AGTGGCAACATAATTTCCACTGCTAGTATTAAGCTCAACTATGGGTGGTGGTCTTCACAT 480  
 QY 481 GCCTATTGTTGTTCAAGACATGCTGTGTTAGCCCTTACTAGGAATCTGGCAGTCCAGCTC 540  
 DB 481 GCCTATTGTTGTTCAAGACATGCTGTGTTAGCCCTTACTAGGAATCTGGCAGTCCAGCTC 540

DB 481 GCCTATTGTTGTTCAAGACATGCTGTGTTAGCCCTTACTAGGAATCTGGCAGTCCAGCTC 540  
 QY 541 GGACAATTTGGCATTAGGGTTAAATGTTTGTCTCCTTTCGGGCTTCTACGGCTTTAGGC 600  
 DB 541 GGACAATTTGGCATTAGGGTTAAATGTTTGTCTCCTTTCGGGCTTCTACGGCTTTAGGC 600  
 QY 601 AAGAAATTTTCAGGAGTAAATAATGAAGAAGAAATTCAGAAATGTAATAAACTTTTCGGGA 660  
 DB 601 AAGAAATTTTCAGGAGTAAATAATGAAGAAGAAATTCAGAAATGTAATAAACTTTTCGGGA 660  
 QY 661 AATTTGAAAGTCCAAAATTTAAATGTTGAGAGTGTGCCAATCGAGCTCTTATCTGGCT 720  
 DB 661 AATTTGAAAGTCCAAAATTTAAATGTTGAGAGTGTGCCAATCGAGCTCTTATCTGGCT 720  
 QY 721 AGTGATGAGGCAAAATAGTGTGAGTGGACACAATCTGTTCAATGATGAGGGTTCACGCTC 780  
 DB 721 AGTGATGAGGCAAAATAGTGTGAGTGGACACAATCTGTTCAATGATGAGGGTTCACGCTC 780  
 QY 781 TGCATTTCTGTAATCAAAAGTGTTCCAATATCCAGATTCT 819  
 DB 781 TGCATTTCTGTAATCAAAAGTGTTCCAATATCCAGATTCT 819  
 RESULT 2  
 AAZ45743  
 ID AAZ45743 standard; cDNA; 816 BP.  
 AC AAZ45743;  
 XX  
 DT 06-APR-2000 (first entry)  
 XX  
 DE Secoisolariciresinol dehydrogenase cDNA clone DEHY130.  
 KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;  
 KW lignan, matairesinol; health-protecting lignan; phytoestrogen;  
 KW enterolactone; enterodiol; nutraceutical; dietary supplement;  
 KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.  
 OS Forsythia intermedia.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..816  
 FT /tag= a  
 FT /product= "secoisolariciresinol dehydrogenase"  
 FT /note= "no termination codon given; partial  
 sequence"  
 XX  
 PN WO955846-A1.  
 XX  
 PD 04-NOV-1999.  
 XX  
 PF 23-APR-1999; 99WO-US08975.  
 XX  
 PR 24-APR-1998; 98US-0082977.  
 XX  
 PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.  
 XX  
 PI Xia Z, Costa MA, Davin LB, Lewis NG;  
 XX  
 XX WPI: 2000-126356/11.  
 DR P-PSDB: AAY54420.  
 XX  
 PT New nucleic acid molecule encoding an enzyme involved in lignan  
 PT biosynthetic pathway, useful for producing large amounts of lignans  
 PS Example 2; Page 61-62; 66pp; English.  
 XX The present sequence encodes a secoisolariciresinol dehydrogenase  
 CC protein. The enzyme is involved in the lignan biosynthetic pathway.  
 CC The secoisolariciresinol dehydrogenase proteins have a molecular  
 CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require  
 CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase  
 CC nucleic acids are used for the recombinant expression of the enzymes.

**QY**      784    AATTCTGTAATCAAAAGTGTTCCAAATATCCAGATTCT    819  
             |  
             |||  
**Dd**      781    AATTCTGTAATCAAAAGTGTTCCAAATATCCAGATTCT    816  
             |  
             |||

**RESULT 3**

AAZ45731 ID # AAZ45731 standard; cDNA; 831 BP.

XX AC AAZ45731;

XX AC

XX XX

DT 06-APR-2000 (first entry)

XX XX

DE Scisolariciresinol dehydrogenase cDNA clone SMDHY321 .

XX XX

KW Scisolariciresinol dehydrogenase; lignan biosynthetic pathway;  
KW lignan; matairesinol; health-protecting lignan; phytoestrogen;  
KW enterolactone; enterodiol; nutraceutical; dietary supplement;  
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.  
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.

xx Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;  
 xx lignan; matairesinol; health-protecting lignan; phytoestrogen;  
 kw enterolactone; enterodiol; neuriceutical; dietary supplement;  
 kw (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.  
 xx  
 OS Forsythia intermedia.

OS Forsythia intermedia.  
XX (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.  
KW entolactone; enterodiol; neuroticeutical; dietary supplement;  
KW lignan; matairesinol; heart-protecting lignan; phytoestrogen;  
KW lignan; matairesinol; heart-protecting lignan; phytoestrogen;

FH	key	Location/Qualifiers
FT	CDS	1..831

```

EI      /"Cag= a
ET      /product= "secoisolaricinesinol dehydrogenase"

```

FT XX  
/note= "no termination codon given"

PN  
XX  
WU9955846-A.I.

FD  
XX  
04-NOV-1995.

PF  
XX  
Z3-APR-1999; 99WU-US08975.

PR 24-APR-1998; 9805-0082977.  
XX

PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.  
XX  
XX

PI XIA Z, Costa MA, Davin LB, Lewis NG;  
XX

DR WPL; 2000-126356/11.  
DR P-PSDB: AAY54413.

XX  
PT  
New nucleic acid molecule encoding an enzyme involved in ligan

FT biosynthetic pathway, useful for producing large amounts of lignans - -  
XX

PS  
XX  
CLAIM 9; Page 48-49; 66pp; English.

CC The present sequence encodes a secoisolariciresinol dehydrogenase  
CC protein. The enzyme is involved in the lignan biosynthetic pathway.

CC The secoisoflavanones have a molecular weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require

CC NAD OF NADP AS A COFACTOR: THE SECONDARY AMINO GROUP OF THE ENZYME.  
CC nucleic acids are used for the recombinant expression of the enzymes.

It is also used to obtain expression of secoisolaricresinol dehydrogenase in plants or to alter lignan

CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase biosynthesis: The enzyme is used for production of the pharmacologically

CC proteins and nucleic acids can be utilized to: elevate or otherwise alter the levels of health-protecting lignans, including phytoestrogens

CC such as enterolactone and enterodiol, in plant species, including vegetables, grains and fruits and to food items incorporating material

derived from such genetically altered plants; genetically altered plants  
 CC species to provide an abundant, natural supply of lignans useful for  
 CC

cc a variety of purposes, for example as nutraceuticals and dietary supplements: to genetically alter living organisms to produce an

abundant supply of optically pure lignans having desirable biological properties, for example (-)-trachelogenin which possesses antiviral

properties, and (-)-podophyllotoxin.  
XX

sequence 831 BP; 244 A; 169 C; 192 G; 220 T; 0 other;  
5Q

```
Query Match      79.0%; Score 647; DB 21; Length 831;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 797; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 20 TCGCAAGAGCGCTAGAGAGAAAGTTGCCCTTATACAGGAGGAGCCAGTGGAAATTTGGAG 79
DB 32 TCGCAAGAGCGCTAGAGAGAAAGTTGCCCTTATACAGGAGGAGCCAGTGGAAATTTGGAG 91
QY 80 AAACACAGCAAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTCATCTCCAAG 139
DB 92 AAACACAGCAAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTCATCTCCAAG 151
QY 140 ATGAATTAGGTCTACTGCTAGTGTGAGGCGCATTTGGCAGCTTCCAAATTCACCTACATCCACT 199
DB 152 ATGAATTAGGTCTACTGCTAGTGTGAGGCGCATTTGGCAGCTTCCAAATTCACCTACATCCACT 211
QY 200 GTGATGTTACTAATGAAGACGGTGTAAANAATGCCGTGGAGCAACACAGTTTCAACCTATG 259
DB 212 GTGATGTTACTAATGAAGACGGTGTAAANAATGCCGTGGAGCAACACAGTTTCAACCTATG 271
QY 260 GAAACTGGACATTATGTTCCAGCAATGCAGGAATTTCTGATCCCAACAGGCCCGCATCA 319
DB 272 GAAACTGGACATTATGTTCCAGCAATGCAGGAATTTCTGATCCCAACAGGCCCGCATCA 331
QY 320 TAGACAACGAAAGACAGACTTTTGAACGGTCTTCAGTGTAAATGTAACCGGAGTTTTC 379
DB 332 TAGACAACGAAAGACAGACTTTTGAACGGTCTTCAGTGTAAATGTAACCGGAGTTTTC 391
QY 380 TATGATGAAGCAGCAGCAGCAGCTTTATGATTCACGACGAGTGCGCAACATAATTTCCA 439
DB 392 TATGATGAAGCAGCAGCAGCAGCTTTATGATTCACGACGAGTGCGCAACATAATTTCCA 451
QY 440 CTGCTAGTTTAACTCAACTATGGTGGTGGTCTTCATGCGCTATTTGGTTCARAAG 499
DB 452 CTGCTAGTTTAACTCAACTATGGTGGTGGTCTTCATGCGCTATTTGGTTCARAAG 511
QY 500 ATGCTGTTAGCCCTTACTAGCAATCGGAGTGGAGCTGGCAATTTGGCAATTTAGG 559
DB 512 ATGCTGTTAGCCCTTACTAGCAATCGGAGTGGAGCTGGCAATTTGGCAATTTAGG 571
QY 560 TTAAATTTGTTCTCTCTTGGGCTTCTACGGCTTTAGGCAAGAAATTTTCAGGGATTA 619
DB 572 TTAAATTTGTTCTCTCTTGGGCTTCTACGGCTTTAGGCAAGAAATTTTCAGGGATTA 631
QY 620 AAATGAAGAAATTTGAGAAATGTAATAAATTTTTCGCGGAAATTTGAAAGGTCCTCAAAAT 679
DB 632 AAATGAAGAAATTTGAGAAATGTAATAAATTTTTCGCGGAAATTTGAAAGGTCCTCAAAAT 691
QY 680 TTAATGTTGAGGATGTTGCCAATGCAAGCTCTTTATCTGGCTAGTGATGAGGCAAAATACG 739
DB 692 TTAATGTTGAGGATGTTGCCAATGCAAGCTCTTTATCTGGCTAGTGATGAGGCAAAATACG 751
QY 740 TGAGTGACACAATCTGTTCAATGATGGAGGTTTCAAGCTCTGCAATTTCTGTAATCAAG 799
DB 752 TGAGTGACACAATCTGTTCAATGATGGAGGTTTCAAGCTCTGCAATTTCTGTAATCAAG 811
QY 800 TGTTCCAATATCCAGATTTCT 819
DB 812 TGTTCCAATATCCAGATTTCT 831

RESULT 4
AAZ45732
ID AAZ45732 standard; cDNA; 819 BP.
AC AAZ45732;
XX AAZ45732;
XX 06-APR-2000 (first entry)
XX Secoisolaricresinol dehydrogenase cDNA clone SMDEHV431.
XX Secoisolaricresinol dehydrogenase; lignan biosynthetic pathway;
XX Lignan; matairesinol; health-protecting lignan; phytoestrogen.
```

```
KW enterolactone; enterodiol; neutriceutical; dietary supplement;
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.
XX Forsythia intermedia.
OS
XX Key Location/Qualifiers
FH CDS 1..819
FT /*Lag= a
FT /product= "secoisolaricresinol dehydrogenase"
FT /note= "no termination codon given"
XX
XX W09555846-A1.
XX
XX 04-NOV-1999.
XX
XX 23-APR-1999; 99WO-US08975.
XX
XX 24-APR-1998; 98US-0082977.
XX
XX (UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX
XX Xia Z, Costa MA, Davin LB, Lewis NG;
XX WPI; 2000-126356/11.
XX P-PSDB; AAY54414.
XX
XX New nucleic acid molecule encoding an enzyme involved in lignan
XX biosynthetic pathway, useful for producing large amounts of lignans
XX
XX Claim 9; Page 50-51; 66pp; English.
XX
XX The present sequence encodes a secoisolaricresinol dehydrogenase
XX protein. The enzyme is involved in the lignan biosynthetic pathway.
XX The secoisolaricresinol dehydrogenase proteins have a molecular
XX weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require
XX NAD or NADP as a cofactor. The secoisolaricresinol dehydrogenase
XX nucleic acids are used for the recombinant expression of the enzymes.
XX It is also used to obtain expression or enhanced expression of
XX secoisolaricresinol dehydrogenase in plants or to alter lignan
XX biosynthesis. The enzyme is used for production of the pharmacologically
XX active lignan, matairesinol. The secoisolaricresinol dehydrogenase
XX proteins and nucleic acids can be utilized to: elevate or otherwise
XX alter the levels of health-protecting lignans, including phytoestrogens
XX such as enterolactone and enterodiol, in plant species, including
XX vegetables, grains and fruits and to food items incorporating material
XX derived from such genetically altered plants; genetically alter plant
XX species to provide an abundant, natural supply of lignans useful for
XX a variety of purposes, for example as nutraceuticals and dietary
XX supplements; to genetically alter living organisms to produce an
XX abundant supply of optically pure lignans having desirable biological
XX properties, for example (-)-trachelogenin which possesses antiviral
XX properties, and (-)-podophyllotoxin.
XX
XX Sequence 819 BP; 235 A; 171 C; 191 G; 220 T; 2 other;
XX
Query Match      11.6%; Score 95; DB 21; Length 819;
Best Local Similarity 99.0%; Pred. No. 3.8e-38;
Matches 195; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 84 CACAGCAAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGATGCCAAGATGA 143
DB 84 CACAGCAAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGATGCCAAGATGA 143
QY 144 ATTAGGTCACTCAGTTGTGCGAGGCCATTGGCAGCTTCCAAATTCACCTACATCCACTGTGA 203
DB 144 ATTAGGTCACTCAGTTGTGCGAGGCCATTGGCAGCTTCCAAATTCACCTACATCCACTGTGA 203
QY 204 TGTTCCTAATGAAGACGGTGTAAANAATGCCGTGGAGCAACACAGTTTCAACCTATGAAA 263
DB 204 TGTTCCTAATGAAGACGGTGTAAANAATGCCGTGGAGCAACACAGTTTCAACCTATGAAA 263
QY 264 ACTGCACATTATGTTCA 280
DB 264 ACTGCACATTATGTTCA 280
```



```
Db 264 ACTGGACATTATGTTCA 280
RESULT 5
ID AA245734 standard; cDNA; 828 BP.
XX
AC AA245734;
XX
DT 06-APR-2000 (first entry)
XX
DE Secoisolariciresinol dehydrogenase cDNA clone SMDEHY631.
XX
KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;
lignan; matairesinol; health-protecting lignan; phytoestrogen;
KW enterolactone; enterodiol; neutriceutical; dietary supplement;
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.
XX
OS Forsythia intermedia.
XX
FH Key Location/Qualifiers
FT CDS 1..828
FT /tag= a
FT /product= "secoisolariciresinol dehydrogenase"
FT /note= "no termination codon given"
XX
PN WO9955846-A1.
XX
PD 04-NOV-1999.
XX
PF 23-APR-1999; 99WO-US089975.
XX
PR 24-APR-1998; 98US-0082977.
XX
PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX
PI Xia Z, Costa MA, Davin LB, Lewis NG;
XX
DR WPI; 2000-126356/11.
DR P-PSDB; AAY54416.
XX
PT New nucleic acid molecule encoding an enzyme involved in lignan
biosynthetic pathway, useful for producing large amounts of lignans
XX
PS Claim 9; Page 55-56; 66pp; English.
XX
CC The present sequence encodes a secoisolariciresinol dehydrogenase
protein. The enzyme is involved in the lignan biosynthetic pathway.
CC The secoisolariciresinol dehydrogenase proteins have a molecular
weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require
CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase
nucleic acids are used for the recombinant expression of the enzymes.
CC It is also used to obtain expression or enhanced expression of
secoisolariciresinol dehydrogenase in plants or to alter lignan
biosynthesis. The enzyme is used for production of the pharmacologically
active lignan, matairesinol. The secoisolariciresinol dehydrogenase
proteins and nucleic acids can be utilized to: elevate or otherwise
alter the levels of health-protecting lignans, including phytoestrogens
such as enterolactone and enterodiol, in plant species, including
vegetables, grains and fruits and to food items incorporating material
derived from such genetically altered plants; genetically alter plant
species to provide an abundant, natural supply of lignans useful for
a variety of purposes, for example as neutriceuticals and dietary
supplements; to genetically alter living organisms to produce an
abundant supply of optically pure lignans having desirable biological
properties, for example (-)-trachelogenin which possesses antiviral
properties, and (-)-podophyllotoxin.
XX
SQ Sequence 828 BP; 235 A; 172 C; 196 G; 225 T; 0 other;
Query Match 11.1%; Score 91; DB 21; Length 828;
Best Local Similarity 100.0%; Pred. No. 4.1e-36;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 CACAGCAAACTCTTCTCCCAACATGAGCCAAAGTTGCCATTGCTGTATGTCACAGATGA 143
|||||
Db 93 CACAGCAAACTCTTCTCCCAACATGAGCCAAAGTTGCCATTGCTGTATGTCACAGATGA 152
|||||
QY 144 ATTAGTCACTCAGTTGTCGAGGCCATTGGC 174
|||||
Db 153 ATTAGTCACTCAGTTGTCGAGGCCATTGGC 183
|||||

RESULT 6
AAL38432
ID AAL38432 standard; DNA; 4505 BP.
XX
AC AAL38432;
XX
DT 15-AUG-2002 (first entry)
XX
DE Plasmid pAPI249 encoding secoisolariciresinol dehydrogenase.
XX
KW Guaiacyl (G)-lignan; monocotyledon plant; rice; food additive;
seed-specific transcriptional regulatory region; dehydrogenase; gene; ds.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 953..1783
FT /tag= a
FT /product= "Protein of secoisolariciresinol dehydrogenase"
FT /note= "b"
XX
PN WO200220548-A1.
XX
PD 14-MAR-2002.
XX
PF 04-SEP-2001; 2001WO-US27500.
XX
PR 07-SEP-2000; 2000US-230632P.
XX
PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.
PA (PHYT-) APPLIED PHYTOLOGICS INC.
XX
PI Lewis NG, Davin LB, Huang N;
XX
DR WPI; 2002-425767/45.
DR P-PSDB; AAO21494, AAO21498.
XX
PT Increasing guaiacyl-lignan content in monocotyledon plants, by
transforming plant with chimeric gene construct having seed-specific
transcriptional regulator linked to gene encoding protein involved in
G-lignan formation
XX
PS Claim 9; Fig 10; 136pp; English.
XX
CC The invention relates to a method for increasing the guaiacyl (G)-lignan
content in seeds of a monocotyledon plant, comprising selecting at least
one protein or enzyme integral to the pathway leading to G-lignan
formation, stably transforming a monocotyledon plant with chimeric gene
(CG) constructs having a seed-specific transcriptional regulatory region
operably linked to a nucleic acid sequence encoding the enzyme. The
method of the invention is useful for stably transforming a
monocotyledonous plant (e.g. rice) with CG constructs resulting in
increased expression of the genes encoded by CG constructs. The G-lignan
enriched seed composition is useful as a food additive. This
polynucleotide sequence represents plasmid pAPI249 encoding
secoisolariciresinol dehydrogenase relating to the invention.
XX
SQ Sequence 4505 BP; 1166 A; 1096 C; 1099 G; 1144 T; 0 other;
Query Match 11.1%; Score 91; DB 24; Length 4505;
Best Local Similarity 100.0%; Pred. No. 4.2e-36;
```

Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CACAGCAAAACTCTTCTCCACATGGAGCCAAAGTTGCCATTGCTGATGTCACAGATGA 143  
|||||  
Db 1045 CACAGCAAAACTCTTCTCCACATGGAGCCAAAGTTGCCATTGCTGATGTCACAGATGA 1104  
|||||

QY 144 ATTAGGTCACCTAGTTGTCGAGGCCATTGGC 174  
|||||  
Db 1105 ATTAGGTCACCTAGTTGTCGAGGCCATTGGC 1135  
|||||

RESULT 7  
AAZ45733  
ID AAZ45733 standard; cDNA; 831 BP.  
XX AC  
XX AAZ45733;  
XX  
DT 06-APR-2000 (first entry)  
XX  
DE Secoisolariciresinol dehydrogenase cDNA clone SMDEHY511.  
XX  
KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;  
KW lignan; matairesinol; health-protecting lignan; phytoestrogen;  
KW enterolactone; enterodiol; nutraceutical; dietary supplement;  
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.  
XX  
OS Forsythia intermedia.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..831  
FT */\*tag= a*  
FT */product= "secoisolariciresinol dehydrogenase"*  
FT */note= "no termination codon given"*  
XX  
FN WO9955846-A1.  
XX  
PD 04-NOV-1999.  
XX  
PF 23-APR-1999; 99WO-US08975.  
XX  
PR 24-APR-1998; 98US-0082977.  
XX  
PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.  
XX  
PI Xia Z, Costa MA, Davin LB, Lewis NG;  
XX  
DR WPI: 2000-126356/11.  
DR P-PSDB: AAY34415.  
XX  
PI New nucleic acid molecule encoding an enzyme involved in lignan  
biosynthetic pathway, useful for producing large amounts of lignans -  
XX  
PS Claim 9; Page 52-54; 66pp; English.  
XX  
XX The present sequence encodes a secoisolariciresinol dehydrogenase  
protein. The enzyme is involved in the lignan biosynthetic pathway.  
XX The secoisolariciresinol dehydrogenase proteins have a molecular  
weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require  
XX NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase  
nucleic acids are used for the recombinant expression of the enzymes.  
XX It is also used to obtain expression or enhanced expression of  
secoisolariciresinol dehydrogenase in plants or to alter lignan  
XX biosynthesis. The enzyme is used for production of the pharmacologically  
active lignan, matairesinol. The secoisolariciresinol dehydrogenase  
XX proteins and nucleic acids can be utilized to: elevate or otherwise  
alter the levels of health-protecting lignans; including phytoestrogens  
XX such as enterolactone and enterodiol, in plant species, including  
vegetables, grains and fruits and to food items incorporating material  
XX derived from such genetically altered plants; genetically alter plant  
species to provide an abundant, natural supply of lignans useful for  
XX a variety of purposes, for example as nutraceuticals and dietary  
supplements; to genetically alter living organisms to produce an  
XX abundant supply of optically pure lignans having desirable biological

CC properties, for example (-)-trachelogenin which possesses antiviral  
CC properties, and (-)-podophyllotoxin.  
XX  
SQ Sequence 831 BP; 244 A; 181 C; 195 G; 211 T; 0 other;  
Query Match 10.6%; Score 87; DB 21; Length 831;  
Best Local Similarity 100.0%; Pred. No. 4.6e-34;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CACAGCAAAACTCTTCTCCACATGGAGCCAAAGTTGCCATTGCTGATGTCACAGATGA 143  
|||||  
Db 96 CACAGCAAAACTCTTCTCCACATGGAGCCAAAGTTGCCATTGCTGATGTCACAGATGA 155  
|||||

QY 144 ATTAGGTCACCTAGTTGTCGAGGCCCAT 170  
|||||  
Db 156 ATTAGGTCACCTAGTTGTCGAGGCCCAT 182  
|||||

RESULT 8  
AAZ45744  
ID AAZ45744 standard; DNA; 33 BP.  
XX AC AAZ45744;  
XX  
DT 06-APR-2000 (first entry)  
XX  
DE PCR primer DEHY130NTERM for secoisolariciresinol dehydrogenase cDNA.  
XX  
KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;  
KW lignan; matairesinol; health-protecting lignan; phytoestrogen;  
KW enterolactone; enterodiol; nutraceutical; dietary supplement;  
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; PCR primer; ss.  
XX  
OS Forsythia intermedia.  
XX  
FN WO9955846-A1.  
XX  
PD 04-NOV-1999.  
XX  
PF 23-APR-1999; 99WO-US08975.  
XX  
PR 24-APR-1998; 98US-0082977.  
XX  
PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.  
XX  
PI Xia Z, Costa MA, Davin LB, Lewis NG;  
XX  
DR WPI: 2000-126356/11.  
XX  
PI New nucleic acid molecule encoding an enzyme involved in lignan  
biosynthetic pathway, useful for producing large amounts of lignans -  
XX  
PS Example 2; Page 63; 66pp; English.  
XX  
XX PCR primers AAZ45744-45 were used to amplify the secoisolariciresinol  
dehydrogenase cDNA sequence. The enzyme is involved in the lignan  
XX biosynthetic pathway. The secoisolariciresinol dehydrogenase proteins  
have a molecular weight of 27-31 kDa, an isoelectric point of 5.9-6.85,  
XX and require NAD or NADP as a cofactor. The secoisolariciresinol  
dehydrogenase nucleic acids are used for the recombinant expression of  
XX the enzymes. It is also used to obtain expression or enhanced expression of  
secoisolariciresinol dehydrogenase in plants or to alter lignan  
XX biosynthesis. The enzyme is used for production of the pharmacologically  
active lignan, matairesinol. The secoisolariciresinol dehydrogenase  
XX proteins and nucleic acids can be utilized to: elevate or otherwise  
alter the levels of health-protecting lignans; including phytoestrogens  
XX such as enterolactone and enterodiol, in plant species, including  
vegetables, grains and fruits and to food items incorporating material  
XX derived from such genetically altered plants; genetically alter plant  
species to provide an abundant, natural supply of lignans useful for  
XX a variety of purposes, for example as nutraceuticals and dietary  
supplements; to genetically alter living organisms to produce an  
XX abundant supply of optically pure lignans having desirable biological

CC properties, for example (-)-trachelogenin which possesses antiviral  
CC properties, and (-)-podophyllotoxin.

SQ Sequence 33 BP; 11 A; 8 C; 7 G; 7 T; 0 other;

Query Match 3.5%; Score 29; DB 21; Length 33;  
Best Local Similarity 100.0%; Pred. No. 0.00018;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTTCGAACCTGATTCGCAAGAAG 29  
|||||  
Db 5 ATGCAGCTTCGAACCTGATTCGCAAGAAG 33

## RESULT 9

AAZ45741  
ID AAZ45741 standard; DNA; 23 BP.

XX  
AC AAZ45741;

XX 06-APR-2000 (first entry)

DE PCR primer DEHYNTERM1 for secoisolariciresinol dehydrogenase cDNA.

XX Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;  
KW lignan; matairesinol; health-protecting lignan; phytoestrogen;  
KW enterolactone; enterodiol; neutraceutical; dietary supplement;  
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; PCR primer; ss.

XX Forsythia intermedia.

OS  
XX WO9955846-A1.

XX 04-NOV-1999.

XX 23-APR-1999; 99WO-US089975.

XX 24-APR-1999; 98US-0082977.

XX (UNIW ) UNIV WASHINGTON STATE RES FOUND.

XX Xia Z, Costa MA, Davin LB, Lewis NG;

XX WPI; 2000-126356/11.

XX New nucleic acid molecule encoding an enzyme involved in lignan  
PT biosynthetic pathway, useful for producing large amounts of lignans

XX Example 2; Page 60; 66pp; English.

XX The present PCR primer was used to amplify the secoisolariciresinol  
CC dehydrogenase cDNA sequence. The enzyme is involved in the lignan  
CC biosynthetic pathway. The secoisolariciresinol dehydrogenase proteins  
CC have a molecular weight of 27-31 kDa, an isoelectric point of 5.9-6.85,  
CC and require NAD or NADP as a cofactor. The secoisolariciresinol  
CC dehydrogenase nucleic acids are used for the recombinant expression of  
CC the enzymes. It is also used to obtain expression or enhanced expression  
CC of secoisolariciresinol dehydrogenase in plants or to alter lignan  
CC biosynthesis. The enzyme is used for production of the pharmacologically  
CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase  
CC proteins and nucleic acids can be utilized to: elevate or otherwise  
CC alter the levels of health-protecting lignans, including phytoestrogens  
CC such as enterolactone and enterodiol, in plant species, including  
CC vegetables, grains and fruits and to food items incorporating material  
CC derived from such genetically altered plants; genetically alter plant  
CC species to provide an abundant, natural supply of lignans useful for  
CC a variety of purposes, for example as neutraceuticals and dietary  
CC supplements; to genetically alter living organisms to produce an  
CC abundant supply of optically pure lignans having desirable biological  
CC properties, for example (-)-trachelogenin which possesses antiviral  
CC properties, and (-)-podophyllotoxin.

XX Sequence 23 BP; 6 A; 7 C; 5 G; 5 T; 0 other;

Query Match 2.8%; Score 23; DB 21; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.21;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CAGCTTCGAACCTGATTCGCAAG 26  
|||||  
Db 1 CAGCTTCGAACCTGATTCGCAAG 23

## RESULT 10

AAL38441/c  
ID AAL38441 standard; DNA; 21 BP.

XX  
AC AAL38441;

XX 15-AUG-2002 (first entry)

XX Dehydrogenase detecting PCR primer #2.

XX Guaiacyl (G)-lignan; monocotyledon plant; rice; food additive;  
KW seed-specific transcriptional regulatory region; PCR; primer; ss;  
KW dehydrogenase.

XX Unidentified.

XX OS  
XX WO200220548-A1.

XX 14-MAR-2002.

XX 04-SEP-2001; 2001WO-US27500.

XX 07-SEP-2000; 2000US-230632P.

XX (UNIW ) UNIV WASHINGTON STATE RES FOUND.

XX (PHYT-) APPLIED PHYTOLOGICS INC.

XX Lewis NG, Davin LB, Huang N;

XX WPI; 2002-425767/45.

XX Increasing guaiacyl-lignan content in monocotyledon plants, by  
PT transforming plant with chimeric gene construct having seed-specific  
PT transcriptional regulator linked to gene encoding protein involved in  
PT G-lignan formation

XX Example 5; Page 76; 136pp; English.

XX The invention relates to a method for increasing the guaiacyl (G)-lignan  
CC content in seeds of a monocotyledon plant, comprising selecting at least  
CC one protein or enzyme integral to the pathway leading to G-lignan  
CC formation, stably transforming a monocotyledon plant with chimeric gene  
CC (CG) constructs having a seed-specific transcriptional regulatory region  
CC operably linked to a nucleic acid sequence encoding the enzyme. The  
CC method of the invention is useful for stably transforming a  
CC monocotyledonous plant (e.g. rice) with CG constructs resulting in  
CC increased expression of the genes encoded by CG constructs. The G-lignan  
CC enriched seed composition is useful as a food additive. This  
CC polynucleotide sequence represents a dehydrogenase detecting PCR primer  
CC relating to the invention.

XX Sequence 21 BP; 7 A; 8 C; 4 G; 2 T; 0 other;

Query Match 2.6%; Score 21; DB 24; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 GGCTGGTGGTCTTCACATGC 482

Db 21 GGCTGGTGGTCTTCACATGC 1

## RESULT 11

```

AAL38451/C
ID AAL38451 standard; DNA; 21 BP.
AC AAL38451;
XX
XX
DT 15-AUG-2002 (first entry)
DE
XX
XX Gtl dehydrogenase PCR primer #2.
XX
KW Guaiacyl (G)-lignan: monocotyledon plant; rice; food additive;
KW seed-specific transcriptional regulatory region; PCR; primer; ss.
XX
OS Unidentified.
XX
XX WO200220548-A1.
XX
XX
XX 14-MAR-2002.
XX
XX
XX 04-SEP-2001; 2001WO-US27500.
XX
XX
XX 07-SEP-2000; 2000US-230632P.
XX
XX (UNIV ) UNIV WASHINGTON STATE RES FOUND.
XX (PHYT-) APPLIED PHYTOLOGICS INC.
XX
XX Lewis NG, Davin LB, Huang N;
XX
XX WPI; 2002-425767/45.
XX
XX Increasing guaiacyl-lignan content in monocotyledon plants, by
XX transforming plant with chimeric gene construct having seed-specific
XX transcriptional regulator linked to gene encoding protein involved in
XX G-lignan formation
XX
XX Example 5; Page 70; 136pp; English.
XX
XX The invention relates to a method for increasing the guaiacyl (G)-lignan
XX content in seeds of a monocotyledon plant, comprising selecting at least
XX one protein or enzyme integral to the pathway leading to G-lignan
XX formation, stably transforming a monocotyledon plant with chimeric gene
XX (CG) constructs having a seed-specific transcriptional regulatory region
XX operably linked to a nucleic acid sequence encoding the enzyme. The
XX method of the invention is useful for stably transforming a
XX monocotyledonous plant (e.g. rice) with CG constructs resulting in
XX increased expression of the genes encoded by CG constructs. The G-lignan
XX increased seed composition is useful as a food additive. This
XX polynucleotide sequence represents a Gtl dehydrogenase PCR primer
XX relating to the invention.
XX
XX Sequence 21 BP; 7 A; 8 C; 4 G; 2 T; 0 other;
XX
XX
XX Query Match 2.6%; Score 21; DB 24; Length 21;
XX Best Local Similarity 100.0%; Pred. No. 2.2;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 462 GGGTGGTGGTCTTCACATGC 482
DB 21 GGGTGGTGGTCTTCACATGC 1
|||||
RESULT 12
AAFL2470/C
ID AAFL2470 standard; cDNA; 703 BP.
XX
XX AAFL2470;
XX
XX
XX 13-MAR-2001 (first entry)
XX
XX Aspergillus oryzae EST SEQ ID NO:4993.
XX
XX Multiple gene expression: filamentous fungal cell; EST;
XX expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW

```

---

```

KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
XX Aspergillus oryzae.
OS
XX WO200056762-A2.
XX
XX 28-SEP-2000.
XX
XX
XX 22-MAR-2000; 2000WO-US07781.
XX
XX 22-MAR-1999; 99US-0273623.
XX
XX (NOVO ) NOVO NORDISK BIOTECH INC.
XX (NOVO ) NOVO NORDISK AS.
XX
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
XX WPI; 2000-594572/56.
XX
XX Monitoring differential expression of genes in filamentous fungal cells
XX uses fluorescence-labeled nucleic acids isolated from the cells and a
XX substrate of expressed sequence tags
XX
XX Claim 88; Page 2099-2100; 316lpp; English.
XX
XX The present invention describes a method for monitoring differential
XX expression of genes in a first filamentous fungal (FF) cell relative to
XX expression of the same genes in one or more second filamentous fungal
XX cells. The method uses fluorescence-labeled nucleic acids isolated from
XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX are used in the methods for monitoring differential expression of genes
XX in a first filamentous fungal (FF) cell relative to expression of the
XX same genes in one or more second filamentous fungal cells. Monitoring
XX the global expression of genes from FF cells allows the production
XX potential of the microorganisms to be improved. New genes may be
XX discovered, possible functions of unknown open reading frames can be
XX identified and gene copy number variation and stability can be
XX monitored. The expression of genes can be used to study how FF cells
XX adapt to changes in culture conditions, environmental stress, spore
XX morphogenesis, recombination, metabolic or catabolic pathway
XX engineering. Using ESTs provides several advantages over genomic or
XX random DNA clones including elimination of redundancy as one spot on an
XX array equals one gene or open reading frame, and organisation of the
XX microarrays based on function of the gene products to facilitate
XX analysis of the results. AAFL1248 to AAFL1853 represents ESTs from
XX Fusarium venenatum; AAFL1248 to AAFL1853 represents ESTs from Aspergillus
XX niger; AAFL1854 to AAFL14878 represents ESTs from Aspergillus oryzae; and
XX AAFL14879 to AAFL15337 represents ESTs from Trichoderma reesei, which are
XX all specifically claimed in the present invention.
XX
XX Sequence 703 BP; 157 A; 208 C; 189 G; 143 T; 6 other;
XX
XX
XX Query Match 2.3%; Score 19; DB 21; Length 703;
XX Best Local Similarity 100.0%; Pred. No. 23;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 119 TTGCCATTGCTGATGTCCA 137
DB 76 TTGCCATTGCTGATGTCCA 58
|||||
RESULT 13
ABQ68055
ID ABQ68055 standard; DNA; 1029 BP.
XX
XX ABQ68055;
XX
XX
XX 29-AUG-2002 (first entry)
XX
XX Listeria monocytogenes EGD DNA sequence #179.
XX
XX Antibacterial; Listeria; food contamination; mutational analysis;
KW

```

KW infection; ds.  
 XX  
 OS Listeria monocytogenes EGD.  
 XX  
 PN WO200228891-A2.  
 XX  
 XX  
 PD 11-APR-2002.  
 XX  
 XX 04-OCT-2001; 2001WO-FR03061.  
 PF  
 XX 04-OCT-2000; 2000FR-0012697.  
 PR  
 XX  
 XX (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 XX Kunst F, Glaser P;  
 XX  
 XX WPI; 2002-332479/37.  
 DR  
 XX  
 XX New genomic sequences from Listeria species, useful for detection,  
 PT treatment and prevention of infection, also related polypeptides,  
 PT antibodies and modulators -  
 XX  
 XX Claim 16; SEQ ID 868; 180pp; French.  
 PS  
 XX  
 CC The present invention relates to nucleic acid sequences  
 CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes  
 CC and primers for identification and/or detection of Listeria (e.g. as  
 CC contaminants in foods, or mutational analysis) and for analysis of  
 CC gene expression. Proteins encoded by the nucleic acid sequences can be  
 CC used to screen for compounds that modulate gene expression, replication  
 CC and pathogenicity of Listeria (potential therapeutic agents), also for  
 CC treating infections by Listeria, and are useful as immunogens in  
 CC anti-Listeria vaccines.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 1029 BP; 377 A; 164 C; 214 G; 274 T; 0 other;  
 SQ  
 Query Match 2.3%; Score 19; DB 24; Length 1029;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 617 TTAAAAATGAAGAAGATT 635  
 Db 275 TTAAAAATGAAGAAGATT 293  
 |||||  
 RESULT 14  
 ABQ69995  
 ID ABQ69995 standard; DNA; 1029 BP.  
 XX  
 XX AC ABQ69995;  
 AC  
 XX 29-AUG-2002 (first entry)  
 DT  
 XX Listeria monocytogenes EGDe DNA sequence #207.  
 DE  
 XX Antibacterial; Listeria; food contamination; mutational analysis;  
 KW infection; ds.  
 XX  
 XX Listeria monocytogenes EGDe.  
 OS  
 XX  
 XX WO200228891-A2.  
 PN  
 XX  
 XX 11-APR-2002.  
 PD  
 XX  
 XX 04-OCT-2001; 2001WO-FR03061.  
 PF  
 XX  
 XX 04-OCT-2000; 2000FR-0012697.  
 PR  
 XX  
 XX (INSP ) INST PASTEUR.  
 PA

PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 XX Kunst F, Glaser P;  
 XX  
 XX WPI; 2002-332479/37.  
 DR  
 XX  
 XX New genomic sequences from Listeria species, useful for detection,  
 PT treatment and prevention of infection, also related polypeptides,  
 PT antibodies and modulators -  
 XX  
 XX Claim 16; SEQ ID 2808; 180pp; French.  
 PS  
 XX  
 CC The present invention relates to nucleic acid sequences  
 CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes  
 CC and primers for identification and/or detection of Listeria (e.g. as  
 CC contaminants in foods, or mutational analysis) and for analysis of  
 CC gene expression. Proteins encoded by the nucleic acid sequences can be  
 CC used to screen for compounds that modulate gene expression, replication  
 CC and pathogenicity of Listeria (potential therapeutic agents), also for  
 CC treating infections by Listeria, and are useful as immunogens in  
 CC anti-Listeria vaccines.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 1029 BP; 377 A; 164 C; 214 G; 274 T; 0 other;  
 SQ  
 Query Match 2.3%; Score 19; DB 24; Length 1029;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 617 TTAAAAATGAAGAAGATT 635  
 Db 275 TTAAAAATGAAGAAGATT 293  
 |||||  
 RESULT 15  
 ABL78811  
 ID ABL78811 standard; cDNA; 480 BP.  
 XX  
 XX AC ABL78811;  
 AC  
 XX 17-MAY-2002 (first entry)  
 DT  
 XX Human ovarian cancer related cDNA clone SEQ ID NO:1789.  
 DE  
 XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.  
 KW  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX WO200192581-A2.  
 PN  
 XX  
 XX 06-DEC-2001.  
 PD  
 XX  
 XX 29-MAY-2001; 2001WO-US17756.  
 PF  
 XX  
 XX 26-MAY-2000; 2000US-207484P.  
 PR  
 XX  
 XX (CORI-) CORIXA CORP.  
 PA  
 XX  
 XX Algate PA, Harlocker SL, Jones R;  
 PI  
 XX  
 XX WPI; 2002-122075/16.  
 DR  
 XX  
 XX Composition for therapy and diagnosis of ovarian cancer comprising  
 PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding  
 PT polypeptide, antibody specific to polypeptide or T cell expressing  
 PT polypeptide -  
 XX  
 XX Claim 1; SEQ ID 1789; 489pp; English.  
 PS  
 XX  
 CC The present invention describes a composition (I) comprising: carriers  
 CC and immunostimulants; and a polypeptide (II) of a ovarian tumour  
 CC

CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence  
 CC (SI) from the 10912 nucleotide sequences as given in ABL77023 to  
 CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell  
 CC population of (II), or antigen presenting cells that express (II).  
 CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to  
 CC (SI) can be used for detecting ovarian cancer in a patient's biological  
 CC sample preferably serum or ovarian tissue. The method comprises  
 CC contacting a biological sample from a patient with (IV), detecting the  
 CC amount of polynucleotide hybridising to (IV) and comparing the amount to  
 CC a predetermined cutoff value and thereby detecting ovarian cancer in the  
 CC patient, where the amount of polynucleotide hybridising to (IV) is  
 CC detected preferably by polymerase chain reaction (PCR). (I) comprising  
 CC (III) and/or (II) is useful for stimulating and/or expanding T cells  
 CC specific for an ovarian tumour protein comprising contacting T cells  
 CC with (III) or (II). (III) is useful in design and preparation of  
 CC ribozyme molecules for inhibiting expression of the tumour polypeptides  
 CC and proteins in tumour cells; and to isolate a full length gene from a  
 CC suitable library e.g., a tumour cDNA library using well known  
 CC techniques.

XX  
 SQ Sequence 480 BP; 133 A; 98 C; 91 G; 157 T; 1 other;

Query Match 2.2%; Score 18; DB 24; Length 480;  
 Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGAGGAGCCAGTGGGATT 75  
 Db 424 GGAGGAGCCAGTGGGATT 441

Search completed: February 18, 2003, 18:37:14  
 Job time : 255 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 17:47:44 ; Search time 2348 Seconds  
(without alignments)  
10151.277 Million cell updates/sec

Title: US-09-673-918a-1

Perfect score: 819

Sequence: 1 atgcagcttcgaactgcatt.....tgtccaataatccagattctt 819

Scoring table: **OLIGO\_NUC**  
Gapop 60.0, Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rtd.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	647	79.0	831	8	AF352735	AF352735 Forsythia
2	22	2.7	15072	1	AE006989	AE006989 Mycobacte
3	22	2.7	38970	1	MTCY10G2	Z95339 Mycobacteri
4	21	2.6	102655	2	AC099466	AC099466 Rattus no
c 5	21	2.6	119807	2	AP003753	AP003753 Oryza sat
c 6	21	2.6	181600	2	AC015813	AC015813 Homo sapi
c 7	21	2.6	202361	2	AC094900	AC094900 Rattus no
c 8	20	2.4	963	3	AF218330	AF218330 Leontodon
c 9	20	2.4	26758	3	U23181	U23181 Caenorhabdi
10	20	2.4	52994	9	AL137004	AL137004 Human DNA
c 11	20	2.4	95315	2	AC098131	AC098131 Rattus no
c 12	20	2.4	150877	2	AC104680	AC104680 Sus scrof
c 13	20	2.4	176699	2	AC021284	AC021284 Homo sapi
14	20	2.4	180748	2	AC129570	AC129570 Mus muscu
15	20	2.4	181210	9	AC011742	AC011742 Homo sapi
16	20	2.4	187508	9	AC104078	AC104078 Homo sapi
17	20	2.4	199274	2	AC124006	AC124006 Mus muscu
c 18	20	2.4	210821	10	AL731814	AL731814 Mouse DNA
c 19	20	2.4	211672	2	AC124694	AC124694 Mus muscu
c 20	20	2.4	221617	2	AC122394	AC122394 Mus muscu
21	19	2.3	702	11	PM7G8G	AL685209 Penicilli
22	19	2.3	783	8	AF092601	AF092601 Munozia
23	19	2.3	783	8	AF092602	AF092602 Liabum gl
24	19	2.3	783	8	AF092603	AF092603 Sinclairi
25	19	2.3	906	8	AF218343	AF218343 Picris pa
26	19	2.3	915	8	AF218351	AF218351 Hieracium
27	19	2.3	939	8	AF218346	AF218346 Taraxacum
28	19	2.3	949	8	AF218336	AF218336 Pilosella
29	19	2.3	960	8	AF218325	AF218325 Tolpis vi
30	19	2.3	961	8	AF218327	AF218327 Tolpis az
31	19	2.3	961	8	AF218329	AF218329 Tolpis co
32	19	2.3	966	8	AF218350	AF218350 Reichardi
33	19	2.3	967	8	AF218328	AF218328 Tolpis fa
34	19	2.3	969	8	AF218352	AF218352 Urospermu
35	19	2.3	971	8	AF218344	AF218344 Hyoseris
36	19	2.3	973	8	AF218338	AF218338 Andryala
37	19	2.3	980	8	AF218342	AF218342 Tolpis ca
38	19	2.3	987	8	AF218326	AF218326 Tolpis ba
39	19	2.3	1026	8	AF405250	AF405250 Nicotieti
40	19	2.3	1026	8	AF405251	AF405251 Hydropect
41	19	2.3	1026	8	AF405252	AF405252 Adenopapp
42	19	2.3	1026	8	AF405254	AF405254 Chrysacti
43	19	2.3	1026	8	AF405256	AF405256 Boeberoid
44	19	2.3	1026	8	AF405257	AF405257 Dysodiops
45	19	2.3	1026	8	AF405258	AF405258 Dysodia

ALIGNMENTS

RESULT 1  
AF352735

LOCUS  
DEFINITION

Forsythia x intermedia stem secoisolariciresinol dehydrogenase  
831 bp mRNA linear PLN 23-APR-2001

ACCESSION  
VERSION

AF352735  
AF352735.1 GI:13752457

KEYWORDS  
SOURCE

MRNA, partial cds.  
Forsythia x intermedia

ORGANISM

Forsythia x intermedia

REFERENCE

1 (bases 1 to 831)  
Xia,Z.Q., Costa,M.A., Pelissier,H.C., Davin,L.B. and Lewis,N.G.

TITLE	Secoisolaricresinol dehydrogenase purification, cloning, and functional expression. Implications for human health protection
JOURNAL	J. Biol. Chem. 276 (16), 12614-12623 (2001)
MEDLINE	21201084
PUBMED	11278426
REFERENCE	2 (bases 1 to 831)
AUTHORS	Xia,Z.O., Costa,M.A., Davlin,L.B. and Lewis,N.G.
TITLE	Direct Submission
JOURNAL	Submitted (26-FEB-2001) Institute of Biological Chemistry, Washington State University, 461 Clark Hall, Pullman, WA 99164-6340, USA
FEATURES	Location/Qualifiers
source	1..831
	/organism="Foxo3" taxid="55183"
CDS	/db_xref="taxon:55183"
	1..>831
	/note="SDH_F1321; NAD-dependent; catalyzes enantiospecific conversion of (-)-secoisolaricresinol into (-)-matairesinol"
	/codon_start=1
	/product="stem secoisolaricresinol dehydrogenase"
	/protein_id="AAK38665.1"
	/db_xref="GI:13752458"
	/translation="MAATSOVLTATABLEKVALITGGAGIGETTAKLFSOHGAKV AIADVDLGHVSVEAIGTQSNSTYVHCVDYEDGVKNVDYVSTYKGLDIFMSNAGI SDNRPRIINEKADFERVFSVNVYGVFLCHKHAARVIMPARSNILISTASLSTMGK GSSHATGCKSHVGLRVLNLAELGPGFVRLVNCSPFGLPTAUGKSGKLNKEEEN VFINATGKLGKPFNEVDVANAALYLASDEAKYVSGHNLFIIDGGFVSCVNSIKRVFYQPD S"
BASE COUNT	244 a 169 c 192 g 226 t
ORIGIN	
Query Match	79.0%; Score 647; DB 8; Length 831;
Best local Similarity	99.6%; Pred. No. 0;
Matches 797; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
QY	20 TCACAAGAGGCTAGAGAAGAAAGTTCGCCCTTATAACAGGAGGAGCGCAGTGGAAATTGGAG 79
Db	32 TCACAAGAGGCTAGAGAAGAAAGTTCGCCCTTATAACAGGAGGAGCGCAGTGGAAATTGGAG 91
QY	80 AAACCACAGCAAACTCTTCCCAACATGGAGCCAAAGTTCGCAATTCGTGATGTCCAAAG 139
Db	92 AAACCACAGCAAACTCTTCCCAACATGGAGCCAAAGTTCGCAATTCGTGATGTCCAAAG 151
QY	140 ATGAATTAGGTGACATTCAGTTGTCGAGGCCATTCGCACTTCCCAATTCACACTACATCCACT 199
Db	152 ATGAATTAGGTGACATTCAGTTGTCGAGGCCATTCGCACTTCCCAATTCACACTACATCCACT 211
QY	200 GTGATGTTACTAATGAAGACGGCTGTAAAAAATGCCGTGGACAACACAGTTTCAACCTATG 259
Db	212 GTGATGTTACTAATGAAGACGGCTGTAAAAAATGCCGTGGACAACACAGTTTCAACCTATG 271
QY	260 GAAACTGGACATTATGTTACAGCAATCCAGCAATTCCTGATCCCAACAGGCCCGCATCA 319
Db	272 GAAACTGGACATTATGTTACAGCAATCCAGCAATTCCTGATCCCAACAGGCCCGCATCA 331
QY	320 TAGCAACGARAAGACGACATTTCAACGCGTTCTCAGTGTAAATGTACCCGGAGTTTCC 379
Db	332 TAGCAACGARAAGACGACATTTCAACGCGTTCTCAGTGTAAATGTACCCGGAGTTTCC 391
QY	380 TATGCATGAAGCAGCAGCAGCTGTTATGATTCACAGCGCAGTGGCAACATTAATTCCA 439
Db	392 TATGCATGAAGCAGCAGCAGCTGTTATGATTCACAGCGCAGTGGCAACATTAATTCCA 451
QY	440 CTGCTAGTTTAAAGCTCAACTATGGTGGTGGTCTTTTCACATGCCCTATTGTGGTTCAAGC 499
Db	452 CTGCTAGTTTAAAGCTCAACTATGGTGGTGGTCTTTTCACATGCCCTATTGTGGTTCAAGC 511
QY	500 ATGCTGTGTTAGCCCTTACTAGGAATCTGCAGTCGAGCTCGGCAAAATTTGGCATTAGGG 559
Db	512 ATGCTGTGTTAGCCCTTACTAGGAATCTGCAGTCGAGCTCGGCAAAATTTGGCATTAGGG 571
QY	560 TTAATTGTTGTGCTTCCTTTCCGGGCTTCCTACGGCTTTTAGGCAAGAAATTTTCAGGGATTA 619



CDS	complement(550..1236) /gene="MT1065" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK45316.1" /db_xref="GI:13880638" /translation="MPHPTTLMKLTTRCGSAAIDGINEALLAKAAEAKILGTNRIRAD TTVANVSYPTDLDGLAKAMRIIAATKRIQAAGCAVTRVGRDRRAHAAVAA KLRSRAELGRDARAVALRFTGELAELOAAQAEAGQLLDNAKQAVLRAKAKAAALAA RGERDAVAGRCGGLVRVNDLTLLNATRIQVATQRVAGITSDGASRRVSLHDGD ARPDHGSRAR"	
gene	complement(1719..2003) /gene="MT1066" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45317.1" /db_xref="GI:13880639" /translation="MTINYQFGVDVDAHGMIRAOAGSLEAEHQAIISDLVTASDFWGG AGSAACQGFITOLGRNFQVIYEANAHGOKVQAAGNNMAQTDSAVGSSWA"	
CDS	complement(2030..2356) /gene="MT1067" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45318.1" /db_xref="GI:13880640" /translation="MTTSRQGDSSMASRFTWDPHAMRDWAGRFVHAQTVDEARRWW ASAQNSIGAGWGMGAETSLDTMTQNNQAFRINVLNMLHGVDRGLVNDANNYEQQEQAS QQILSS"	
gene	complement(2472..3647) /gene="MT1068" /note="similar to GB:AL123456; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="PPE family protein" /protein_id="AAK45319.1" /db_xref="GI:13880641"	
CDS	complement(2472..3647) /gene="MT1068" /note="similar to GB:AL123456; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="protease-related protein" /protein_id="AAK45324.1" /db_xref="GI:13880646" /translation="MCAHQFFGLVHNPNVAAAIKGRPPPPVDSIDGLPTTVPEPWSV ADFSRYLSTLGIPLPAAGDAVTLHRISSMERAGLLPLGWDPRPLPVNGQYIISOGATISK GORGGNLSTVEFGAELIIPSYNAVTVLAGHDDAGNVDGSMGTGLVVDHNRVITNKH VVTGLAGTSAGLSVSPSSNBAELVNFSGTAPHTLDVAVIKFEMPEGKYIPRLGG MAFRDPDWADEVYVFCYPRVPMTEAMAITVORGEVNVNPAATTIPGRQKIFLXSAIARP GNSGGPIVAQDGRVIGLVVEDSAEAPSTGCTGPNNAFFYKGIPISSSEVIRALDELDFGCI VEMDTLP"	
gene	8229..8852 /gene="MT1074" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK45325.1" /db_xref="GI:13880647" /translation="MCAKPYLIDTIAHMAIDRLVEAAEQHYVTTTDRARDIGVDPV QURLLAGRGRLEVRGVYRVPVLPGEHDDLAAAVSWTLGRGVISHESALAHALAD VNPSTRILHTVPRNNHPRAAGGELYRVHRRDLQAAHVTSVDGIPVTTVARTIKDCVKTG	
CDS	8229..8852 /gene="MT1074" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK45324.1" /db_xref="GI:13880646" /translation="MCAHQFFGLVHNPNVAAAIKGRPPPPVDSIDGLPTTVPEPWSV ADFSRYLSTLGIPLPAAGDAVTLHRISSMERAGLLPLGWDPRPLPVNGQYIISOGATISK GORGGNLSTVEFGAELIIPSYNAVTVLAGHDDAGNVDGSMGTGLVVDHNRVITNKH VVTGLAGTSAGLSVSPSSNBAELVNFSGTAPHTLDVAVIKFEMPEGKYIPRLGG MAFRDPDWADEVYVFCYPRVPMTEAMAITVORGEVNVNPAATTIPGRQKIFLXSAIARP GNSGGPIVAQDGRVIGLVVEDSAEAPSTGCTGPNNAFFYKGIPISSSEVIRALDELDFGCI VEMDTLP"	
gene	8229..8852 /gene="MT1074" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK45325.1" /db_xref="GI:13880647" /translation="MCAKPYLIDTIAHMAIDRLVEAAEQHYVTTTDRARDIGVDPV QURLLAGRGRLEVRGVYRVPVLPGEHDDLAAAVSWTLGRGVISHESALAHALAD VNPSTRILHTVPRNNHPRAAGGELYRVHRRDLQAAHVTSVDGIPVTTVARTIKDCVKTG	

gene  
 TDYQLRAAIEAAEAGTLLRRGSAALRAALDETTAGLRARPKRASA"  
 8849. .9730  
 /gene="M1075"

Query Match 2.7% Score 22; DB 1; Length 15072;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 683 ATGTTGAGGATGTTGCCAATGC 704  
 |||||||  
 Db 4151 ATGTTGAGGATGTTGCCAATGC 4172

RESULT 3  
 MTCY10G2 38970 bp DNA linear BCT 03-AUG-2001  
 Mycobacterium tuberculosis H37Rv complete genome; segment 47/162.  
 ACCESSION Z92539 AL123456  
 VERSION Z92539.1 GI:3261714

KEYWORDS  
 SOURCE Mycobacterium tuberculosis H37Rv.

ORGANISM  
 Mycobacterium tuberculosis H37Rv  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Corynebacterineae; Mycobacteriaceae;  
 Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE  
 AUTHORS Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.,  
 Harris D., Gordon S.V., Eiglmeller K., Gas S., Barry III C.E.,  
 Tekaita F., Badcock K., Basham B., Brown D., Chillingworth T.,  
 Connor R., Davies K., Devlin K., Feltwell T., Gentles S.,  
 Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J.,  
 Moule S., Murphy L., Oliver S., Osborne J., Quail M.A.,  
 Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S.,  
 Squares S., Squires R., Sulston J.E., Taylor K., Whitehead S. and  
 Barrell B.G.

TITLE  
 Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence

JOURNAL  
 Nature 393 (6685), 537-544 (1998)

MEDLINE  
 96295987

PUBMED  
 9634230

REFERENCE  
 2 (bases 1 to 38970)

AUTHORS  
 Parkhill J.

TITLE  
 Direct Submission

JOURNAL  
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium  
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome  
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique  
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,  
 75724 Paris Cedex 15, France E-mail: parkhillsanger.ac.uk  
 On Jun 27, 1998 this sequence version replaced gi:1869979.

COMMENT

Notes:  
 Details of M. tuberculosis sequencing at the Sanger Centre are  
 available on the World Wide Web  
 (URL: [http://www.sanger.ac.uk/Projects/M\\_tuberculosis/](http://www.sanger.ac.uk/Projects/M_tuberculosis/)) CDS have  
 been renumbered from the original cosmid submissions but the old  
 gene designations are in brackets after the new gene numbers.  
 Gene prediction was based on a Hidden Markov Model of TB genes  
 implemented in TBparse (Krogh) supplemented with visual inspection  
 of positional base preference in codons, especially where there is  
 an increase in the observed/expected third position G + C.  
 CAUTION: In some cases we may not have predicted the correct  
 initiation codon. Where possible we choose an initiation codon  
 (atg, gtg, or ttg) which is preceded by an upstream ribosome  
 binding site sequence (optimally 5-13bp before the initiation  
 codon). If this cannot be identified we choose the most upstream  
 initiation codon.

FEATURES

source

1. .38970  
 Location/Qualifiers  
 /organism="Mycobacterium tuberculosis H37Rv"  
 /strain="H37Rv"  
 /db\_xref="taxon:83332"  
 <1. .38452  
 /organism="Mycobacterium tuberculosis H37Rv"  
 /strain="H37Rv"

source

/db\_xref="taxon:83332"  
 /clone="Y10G2"  
 complement(51. .626)  
 /gene="pth"  
 complement(51. .626)  
 /gene="pth"  
 /note="Rv1014c, (MTCY10G2.35), pth, len: 191. Probable  
 peptidyl-trna hydrolase, similar to eg PTH\_ECOLI P23932  
 peptidyl-trna hydrolase (ec 3.1.1.29) (194 aa), fasta  
 scores, opt: 472, E(): 2.3e-25, (39.6% identity in 187 aa  
 overlap)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="pth"  
 /protein\_id="CAB06865.1"  
 /db\_xref="GI:1870014"  
 /db\_xref="SWISS-PROT:P96386"  
 /translation="MAEPLLVVGLGNPGANGYARTRNILGVVADLLAARLCAKFAKHK  
 RSGAEVATGRSAGRSILVLAKPCYMNESGROIGLAKFYSVAPANIIVIHDDLDFRC  
 RIRLKIGGGGGHNGRLSRVVAALGTDKDFVRIGIRPPGRKDPAAFLVFNFTFAERA  
 EVPTICDAAADTELLIEQMEPAQNRVHAH"  
 complement(639. .1286)  
 /gene="rply"  
 complement(639. .1286)  
 /gene="rply"  
 /note="Rv1015c, (MTCY10G2.34), rply, len: 215. Probable  
 50S ribosomal protein 125 similar to RL25\_ECOLI P02426 50s  
 ribosomal protein 125. escherichi (94 aa), fasta scores,  
 opt:182, E(): 2.5e-05, (38.4% identity in 86 aa overlap)  
 and t0CTC\_BACSU P14194 general stress protein etc.  
 bacillus subtilis (203 aa), fasta scores, opt: 260, E():  
 1.4e-09, (28.4%identity in 201 aa overlap)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="rply"  
 /protein\_id="CAB06864.1"  
 /db\_xref="GI:1870013"  
 /db\_xref="SWISS-PROT:P96385"  
 /translation="MAKSASNQLRVTVRTETGKASRRARRAGKIPAVLYGHGAEPQH  
 LELPGHDYAAVLRHSGTNALVLTIDAGKOLATKALHIPIRRTIOHADLLVVRGE  
 KVVVEVSVVGGQAGPDLTVTOETNSIEIEAENALSIPEOLTVSIEGAPGQLTAGQI  
 ALPAGVSLISDPDLLVNVVKAPTAETLEBGEVAGAEAEAAVEAGEAEAAEAGESE"  
 complement(1503. .2183)  
 /gene="lpqT"  
 complement(1503. .2183)  
 /gene="lpqT"  
 /note="Rv1016c, (MTCY10G2.33), len: 226. Probable  
 lipoprotein, similar to several M. tuberculosis  
 hypothetical proteins e.g. Y0H3\_MYCTU P71697, Cy21d4.03c  
 (310 aa), fasta scores; opt: 329 z-score: 442.0 E():  
 2e-17, 32.3% identity in 229 aa overlap; contains PS00013  
 Prokaryotic membrane lipoproteinlipid attachment site"  
 /codon\_start=1  
 /transl\_table=11  
 /product="lpqT"  
 /protein\_id="CAB06863.1"  
 /db\_xref="GI:1870012"  
 /db\_xref="SWISS-PROT:P96384"  
 /translation="MAGRRCPQDSVRPLAVAVATLAMSACVAGKSPDFQSILSTS  
 PTTSAVSTTTEVPVPLWLVESVGTGEPVAPSSLTDLTISPTTPPGWAMKPNIP  
 NTNIAKSGSYPTAMLMVFLHRDFDIAELKHGTADARLSTNFTDSDSTADPNFGP  
 SSMIQSGYDLHGRLHTNRIVFPTGAPPKAKORYLVQLTITSLANEVKAASDEAII  
 AGFVAAK"  
 complement(2094. .2126)  
 /gene="lpqT"  
 /note="PS00013 prokaryotic membrane lipoprotein lipid  
 attachment site"  
 complement(2219. .3199)  
 /gene="prsA"  
 complement(2219. .3199)  
 /gene="prsA"  
 /note="Rv1017c, (MTCY10G2.32), prsA, len: 326. Probable  
 ribose-phosphate pyrophosphokinase, similar to eg

misc\_feature

gene

CDS

KPRS\_ECOLI P08330 ribose-phosphate pyrophosphokinase (314 aa), fasta scores, opt: 826, E(): 0, (43.8% identity in 317 aa overlap): contains PS00103 Purine/pyrimidine phosphoribosyl transferases signature: contains PS00144 Asparaginase / glutaminase active site signature 1" /codon\_start=1 /transl\_table=11 /product="prsa" /protein\_id="CAB06862.1" /db\_xref="GI:1870011" /db\_xref="SPTREMBL:P96383" /translation="MSHWDTNRKMLFAGRAHPELAQVAKELDVHVTQDAREFA NGELVRFHESVRGCDALVQSCPAPVNRWLMQELIMIDALKRGSAKRITAVMPFPY ARQDKHGRREPISARLADLKTAKADRIYVLDHTIQGFFGDPGHRGQNLIT GYRDNYDGNMNVVSPDSGRVRIAEKWADALGGVPLAFIKHTRPRVPNVVSRVV GDVAGTCVLIDDMIDTGTGTAGAVALLNDGAGVVIATTAATHGVLSDDPAARLASC GA REVVTNLTPLIGEDKRFQLTVLSIAPLASTIRAVFNGSVTGLFDGDA" complement(2480..2506) /gene="prsa" /note="PS00144 Asparaginase / glutaminase active site signature 1" complement(2486..2524) /gene="prsa" /note="PS00103 Purine/pyrimidine phosphoribosyl transferases signature" complement(3291..4778) /gene="glmu" complement(3291..4778) /gene="glmu" /note="rv1018c, (MTCY10G2.31), glmu, len: 495. Probable udp-n-acetylglucosamine pyrophosphorylase, similar to eg GCAD\_BACSU P14192 udp-n-acetylglucosamine pyrophosphorylase (EC2.7.7.23) (456 aa), fasta scores, opt: 1150, E(): 0, (40.0% identity in 453 aa overlap)" /codon\_start=1 /transl\_table=11 /product="glmu" /protein\_id="CAB06861.1" /db\_xref="GI:1870010" /db\_xref="SPTREMBL:P96382" /translation="MTFPGDTAVLVLAAGPGTRMSRDPKVLHTLAGRSMLSHLVHAI AKLAPQLVILGHQDHRIAPLNGELADTLGRTIDVALQDPLGTGTHAVLGLSLPDI DVAGNVDTGTPDLPDADTLADLIATHRAVSAVTLTLDLPFGRLRTODHE VMAIEQVDTAPSOIREVNAVYFDIALRSALSSNNAAQOELLTDVIAILR SDQGVTHSHVDVAGVNNRVQALAEISLNRVVAHOLAGVTVVDPATWIDV DTVGRDVIHPGTOLLGTOIGRCVVGPDITLDVAVGDGASVVRTHGSSSSIGDG RAAVGPTYLPRCTALGADGKLGAFVEVKNSTIGTCTKVPHLTYVDADIGYSNIGAS SVFVNYDTSKRRTVGVSHVRGTSDTMFVAPVTIGGATGATVVRDVPVPGALVAS AGPQRNIENWVRKRPSPAPAAQASKRASEMACQQPTPPDADQTP" complement(4794..4865) /product="trna-gln" /note="glnT, trna-Gln, anticodon tlg, length = 72" 5033..5626 /gene="rv1019" /gene="rv1019" /gene="rv1019" /note="rv1019, (MTCY10G2.30c), len: 197. Probable transcriptional regulator, similar to many members of the tetR family e.g. MTCY7D11.18c, (34.4% identity in 189 aa overlap). Helix turn helix motif from aa 27-48 (+5.42 SD)" /codon\_start=1 /transl\_table=11 /product="hypothetical protein Rv1019" /protein\_id="CAB06860.1" /db\_xref="GI:1870009" /db\_xref="SPTREMBL:P96381" /translation="MTGTERRHLQIGIARSLFAERYGDTSTIEEIAQRAVNSKPVVYE HFGKRGYAVVVDREMSALLDGTSSLTNNRSVRVERVALALITYVEERTDGRIM IRDSPASISSTGYSLLNDVAYSQVSSILAGDFARRGLDPLAPLVAQALVGSVSMTAQ

Query Match 2.7%; Score 22; DB 1; Length 38970; Best Local Similarity 100.0%; Pred. No. 0.97; Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 683 ATGTTGAGGATGTTGCCAATGC 704  
|||||  
Db 29694 ATGTTGAGGATGTTGCCAATGC 29715  
|||||

RESULT 4  
AC099466  
LOCUS  
DEFINITION  
AC099466  
AC099466.3 GI:21730043  
HTG: HTGS\_PHASE1.  
SOURCE  
Norway rat.  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 102655)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Correll,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,R., Louissege,H.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojebokan,I., Rolfe,M., Ruiz,S., Savary,G.,  
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,  
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,K., Tang,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 102655)  
Worley,K.C.  
Direct Submission  
Submitted (15-NOV-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 102655)  
Worley,K.C.  
Direct Submission  
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One

COMMENT

Baylor Plaza, Houston, TX 77030, USA

On Jul 11, 2002 this sequence version replaced gi:17974364.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GTW

Center Clone name: CH20-6319

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 53408 bases at least Q40

Consensus quality: 56431 bases at least Q30

Consensus quality: 58711 bases at least Q20

-----

\* NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 52 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence.

\* as soon as it is available and the accession number will

\* be preserved.

\* 1

\* 1125: contig of 1125 bp in length

\* 1126: gap of unknown length

\* 1226: contig of 1038 bp in length

\* 2264: gap of unknown length

\* 2364: contig of 1314 bp in length

\* 3677: gap of unknown length

\* 3778: contig of 1245 bp in length

\* 5022: gap of unknown length

\* 5122: contig of 1312 bp in length

\* 6434: gap of unknown length

\* 6534: contig of 1525 bp in length

\* 8059: gap of unknown length

\* 8159: gap of unknown length

\* 9232: contig of 1073 bp in length

\* 9332: gap of unknown length

\* 9333: contig of 1363 bp in length

\* 10695: contig of 1011 bp in length

\* 10796: gap of unknown length

\* 11806: contig of 1069 bp in length

\* 11807: gap of unknown length

\* 12975: contig of 1139 bp in length

\* 13076: gap of unknown length

\* 14214: contig of 1187 bp in length

\* 14315: gap of unknown length

\* 14315: contig of 1187 bp in length

\* 15501: gap of unknown length

\* 15601: contig of 1019 bp in length

\* 16820: gap of unknown length

\* 16821: contig of 1354 bp in length

\* 18075: gap of unknown length

\* 18174: contig of 1845 bp in length

\* 20019: gap of unknown length

\* 20119: contig of 1555 bp in length

\* 21674: gap of unknown length

\* 21774: contig of 1482 bp in length

\* 23256: gap of unknown length

\* 23356: contig of 1426 bp in length

\* 24782: gap of unknown length

\* 24882: contig of 1899 bp in length

\* 26781: gap of unknown length

\* 26881: contig of 1300 bp in length

\* 28181: gap of unknown length

\* 28281: contig of 1045 bp in length

\* 29326: gap of unknown length

\* 29427: contig of 1251 bp in length

\* 30677: gap of unknown length

\* 30778: contig of 1277 bp in length

\* 32054: contig of 1277 bp in length

\* 32055

\* 32155: gap of unknown length

\* 33520: contig of 1366 bp in length

\* 33521: gap of unknown length

\* 35765: contig of 2145 bp in length

\* 35865: gap of unknown length

\* 35866: contig of 1473 bp in length

\* 37438: gap of unknown length

\* 37439: contig of 1854 bp in length

\* 39292: gap of unknown length

\* 39322: contig of 2389 bp in length

\* 41781: gap of unknown length

\* 41782: contig of 2550 bp in length

\* 44531: gap of unknown length

\* 44532: contig of 1457 bp in length

\* 45988: gap of unknown length

\* 46088: contig of 1406 bp in length

\* 47494: gap of unknown length

\* 47495: contig of 2604 bp in length

\* 50198: gap of unknown length

\* 50199: contig of 1447 bp in length

\* 51745: gap of unknown length

\* 51845: contig of 1546 bp in length

\* 53391: gap of unknown length

\* 53492: contig of 1497 bp in length

\* 54988: gap of unknown length

\* 55088: contig of 1980 bp in length

\* 57068: gap of unknown length

\* 57158: contig of 2768 bp in length

\* 59336: gap of unknown length

\* 60036: contig of 1746 bp in length

\* 61882: gap of unknown length

\* 64005: contig of 2123 bp in length

\* 64105: gap of unknown length

\* 65295: contig of 1190 bp in length

\* 65395: gap of unknown length

\* 67377: contig of 1982 bp in length

\* 67477: gap of unknown length

\* 68685: contig of 1208 bp in length

\* 68785: gap of unknown length

\* 70812: contig of 2027 bp in length

\* 70912: gap of unknown length

\* 73643: contig of 2731 bp in length

\* 73743: gap of unknown length

\* 77074: contig of 3331 bp in length

\* 77174: gap of unknown length

\* 80666: contig of 3492 bp in length

\* 80766: gap of unknown length

\* 83466: contig of 2700 bp in length

\* 83566: gap of unknown length

\* 87526: contig of 3960 bp in length

\* 87626: gap of unknown length

\* 90412: contig of 2786 bp in length

\* 90512: gap of unknown length

\* 94867: contig of 4355 bp in length

\* 94868: gap of unknown length

\* 97861: contig of 2894 bp in length

\* 97862: gap of unknown length

\* 97962: contig of 4694 bp in length.

FEATURES

Location/Qualifiers

1..102655

/organism="Rattus norvegicus"

/db\_xref="taxon:10116"

Query Match 2.6%; Score 21; DB 2; Length 102655;

Best Local Similarity 100.0%; Pred. No. 3.4;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 605 AATTTTCAGGATTAATAATG 625

|||||

Db 67018 AATTTTCAGGATTAATAATG 67038

RESULT 5

```

AP003753/c
LOCUS       AP003753               119807 bp    DNA        linear   HTG 21-MAR-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 7 clone
            OJ1339_B08, *** SEQUENCING IN PROGRESS ***
ACCESSION   AP003753
VERSION     AP003753.1  GI:14422471
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
            clone:OJ1339_B08.
ORGANISM    Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE   1
AUTHORS     Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE       Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
            clone:OJ1339_B08
JOURNAL     Published Only in Database (2001)
REFERENCE   2
AUTHORS     Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE       Direct Submission
JOURNAL     Submitted (13-JUN-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
            The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
COMMENT     NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source           1..119807
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosomes="7"
                     /clone="OJ1339_B08"
BASE COUNT  34992 a 23815 c 24131 g 36815 t      54 others
ORIGIN

Query Match      2.6%; Score 21; DB 2; Length 119807;
Best Local Similarity 100.0%; Pred. NO. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 545 AATTGGCATTAGGTTAATT 565
|||||
Db 18836 AATTGGCATTAGGTTAATT 18816

RESULT 6
AC015813
LOCUS       AC015813               181600 bp    DNA        linear   HTG 06-AUG-2002
DEFINITION  Homo sapiens clone RP11-159D12, *** SEQUENCING IN PROGRESS ***
ACCESSION   AC015813
VERSION     AC015813.11  GI:22122998
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Birren,B., Nusbaum,C. and Lander,E.
TITLE       1 (bases 1 to 181600)
JOURNAL     Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            3 (bases 1 to 181600)
            Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murray,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
            Phukhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zemlek,L., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Aug 6, 2002 this sequence version replaced gi:21591957.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L398
            Center clone name: 159_D_12
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            * 1 19146: contig of 19146 bp in length
            * 19147 19246: gap of 100 bp
            * 19247 181600: contig of 162354 bp in length.
            Location/Qualifiers
            1..181600
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Rp11-159D12"
            /clone_lib="RP11-11 Human Male BAC"
BASE COUNT  50395 a 41057 c 40282 g 49766 t      100 others
ORIGIN

```

REFERENCE  
AUTHORS

2 (bases 1 to 181600)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Balaban,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B., Brown,A., Castle,A., Collangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

## TITLE

## JOURNAL

Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE  
AUTHORS

3 (bases 1 to 181600)

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murray,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phukhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zemlek,L., Zimmer,A. and Zody,M.

## TITLE

## JOURNAL

Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Aug 6, 2002 this sequence version replaced gi:21591957.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

## FEATURES

## source

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L398  
Center clone name: 159\_D\_12  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 19146: contig of 19146 bp in length  
\* 19147 19246: gap of 100 bp  
\* 19247 181600: contig of 162354 bp in length.

## FEATURES

## source

1..181600  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Rp11-159D12"

## BASE COUNT

## ORIGIN

50395 a 41057 c 40282 g 49766 t 100 others

```

Query Match      2.6%; Score 21; DB 2; Length 181600;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 CCATTGGCAGCTTCCAAATTCCTCA 187
|||||CCCCCCCCCCCCCCCCCCCCCCCC
Db 91881 CCATTGGCAGCTTCCAAATTCCTCA 91901

RESULT 7
AC094900/c
LOCUS          AC094900          202361 bp      DNA      linear      HTG 22-JUL-2002
DEFINITION     Rattus norvegicus clone CH230-6J7, *** SEQUENCING IN PROGRESS ***,
88 unordered pieces.
ACCESSION      AC094900
VERSION        AC094900.5      GI:21909431
KEYWORDS       HTG; HTGS_PHASE1
SOURCE         Rattus norvegicus
ORGANISM       Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 202361)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbacia,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Donthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Ehrhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flegg,N., Ford,J., Foster,P., Frantz,P.,
Gabis,A., Gao,J., Garcia,A., Garner,F., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Jackson,L.E.,
Jacobson,B., Jia,X., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissege,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usman,K., Vasquez,L., Vera,V., Villalob,D., Vinson,K., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczky,R., Woodson,C., Worley,K.,
Wu,C., Wu,Y., Wu,X., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 202361)
Worley,K.C.
Direct Submission
TITLE
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 202361)
Worley,K.C.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (22-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 19, 2002 this sequence version replaced gi:21903566.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html.
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GBRE
Center clone name: CH230-6J7
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 124551 bases at least Q40
Consensus quality: 131182 bases at least Q30
Consensus quality: 135595 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 88 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 1043: contig of 1043 bp in length
1044 1143: gap of unknown length
1144 2195: contig of 1052 bp in length
2195 2295: gap of unknown length
2295 3391: contig of 1096 bp in length
3391 3491: gap of unknown length
3491 4697: contig of 1206 bp in length
4697 4797: gap of unknown length
4797 5836: contig of 1039 bp in length
5836 5937: gap of unknown length
5937 7011: contig of 1075 bp in length
7011 7111: gap of unknown length
7111 8165: contig of 1054 bp in length
8165 8266: gap of unknown length
8266 9503: contig of 1238 bp in length
9503 9603: gap of unknown length
9603 10668: contig of 1065 bp in length
10668 10768: gap of unknown length
10768 12186: contig of 1418 bp in length
12186 12286: gap of unknown length
12286 13747: contig of 1461 bp in length
13747 13847: gap of unknown length
13847 15120: contig of 1273 bp in length
15120 15207: gap of unknown length
15207 16607: contig of 1387 bp in length
16607 16707: gap of unknown length
16707 18470: contig of 1763 bp in length
18470 18570: gap of unknown length
18570 19646: contig of 1076 bp in length
19646 21428: contig of 1682 bp in length
21428 21528: gap of unknown length
21528 22681: contig of 1153 bp in length
22681 22781: gap of unknown length
22781 24556: contig of 1775 bp in length
24556 24657: gap of unknown length
24657 26009: contig of 1353 bp in length
26009 26109: gap of unknown length
26109 27291: contig of 1182 bp in length
27291 27391: gap of unknown length
27391 28689: contig of 1298 bp in length
28689 28690: gap of unknown length

```

\* 28790 29866: contig of 1077 bp in length  
\* 29867 gap of unknown length  
\* 31021: contig of 1055 bp in length  
\* 31121: gap of unknown length  
\* 31122 contig of 1357 bp in length  
\* 32479 32578: gap of unknown length  
\* 33579 33982: contig of 1404 bp in length  
\* 33983 34082: gap of unknown length  
\* 34083 35337: contig of 1255 bp in length  
\* 35338 35437: gap of unknown length  
\* 35438 36895: contig of 1458 bp in length  
\* 36896 gap of unknown length  
\* 36996 38416: contig of 1421 bp in length  
\* 38417 38516: gap of unknown length  
\* 38517 40009: contig of 1493 bp in length  
\* 40010 40109: gap of unknown length  
\* 40110 41677: contig of 1568 bp in length  
\* 41678 41777: gap of unknown length  
\* 41778 43476: contig of 1699 bp in length  
\* 43477 43576: gap of unknown length  
\* 43577 44726: contig of 1150 bp in length  
\* 44727 44826: gap of unknown length  
\* 44827 46863: contig of 2037 bp in length  
\* 46864 46963: gap of unknown length  
\* 47970 47969: contig of 1006 bp in length  
\* 48070 48069: gap of unknown length  
\* 49211: contig of 1142 bp in length  
\* 49312 49311: gap of unknown length  
\* 50769 50768: contig of 1458 bp in length  
\* 50770 50869: gap of unknown length  
\* 50870 53404: contig of 2435 bp in length  
\* 53405 53404: gap of unknown length  
\* 53405 55313: contig of 1909 bp in length  
\* 55314 55413: gap of unknown length  
\* 55414 56546: contig of 1133 bp in length  
\* 56547 56846: gap of unknown length  
\* 56847 57656: contig of 1010 bp in length  
\* 57657 57756: gap of unknown length  
\* 57757 59322: contig of 1466 bp in length  
\* 59323 59322: gap of unknown length  
\* 59323 61550: contig of 2228 bp in length  
\* 61551 61550: gap of unknown length  
\* 61551 63277: contig of 1626 bp in length  
\* 63277 63376: gap of unknown length  
\* 63377 64933: contig of 1557 bp in length  
\* 64934 65033: gap of unknown length  
\* 65034 67183: contig of 2150 bp in length  
\* 67184 67283: gap of unknown length  
\* 67284 69479: contig of 2196 bp in length  
\* 69480 69579: gap of unknown length  
\* 69580 71824: contig of 2245 bp in length  
\* 71825 71924: gap of unknown length  
\* 71925 73839: contig of 1915 bp in length  
\* 73840 73939: gap of unknown length  
\* 73940 75368: contig of 1429 bp in length  
\* 75369 75469: gap of unknown length  
\* 75469 77661: contig of 2193 bp in length  
\* 77662 77761: gap of unknown length  
\* 77762 79659: contig of 1898 bp in length  
\* 79660 79759: gap of unknown length  
\* 79760 82318: contig of 2359 bp in length  
\* 82319 84218: gap of unknown length  
\* 84219 84221: contig of 1803 bp in length  
\* 84222 84321: gap of unknown length  
\* 84322 86104: contig of 1783 bp in length

Query Match 2.6%; Score 21; DB 2; Length 202361;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 AATTTTCAGGATTAATAATG 625  
|||||  
DB 96745 AATTTTCAGGATTAATAATG 96725

RESULT 8  
LOCUS AF218330 963 bp DNA linear PLN 12-JUN-2001  
DEFINITION. Leontodon saxatilis NADH dehydrogenase (ndhf) gene, partial cds; chloroplast gene for chloroplast product.  
ACCESSION AF218330  
VERSION AF218330.1 GI:10180318  
KEYWORDS Leontodon saxatilis.  
SOURCE Leontodon saxatilis.  
ORGANISM Chloroplast Leontodon saxatilis  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae; Lactuceae; Leontodon.  
REFERENCE 1 (bases 1 to 963)  
AUTHORS Park,S.-J., Korompai,E.J., Francisco-Ortega,J., Santos-Guerra,A. and Jansen,R.K.  
TITLE Phylogenetic relationships of Tolpis (Asteraceae: Lactuceae) based on ndhf sequences  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 963)  
AUTHORS Jansen,R.K. and Park,S.-J.  
TITLE Direct Submission  
JOURNAL Submitted (22-DEC-1999) Integrative Biology, University of Texas at Austin, 24th at Whitis, Austin, TX 78712, USA  
FEATURES  
source Location/Qualifiers  
1..963  
/organism="Leontodon saxatilis"  
/organelle="plastid:chloroplast"  
/db\_xref="taxon:122540"  
<1..963  
/gene="ndhf"  
/gene="ndhf"  
/codon\_start=1  
/transl\_table=11  
/product="NADH dehydrogenase"  
/protein\_id="AAG14009.1"  
/db\_xref="GI:10180319"  
/translation="NDSWLYSPFIIAMTAGTAFYMCRIYLLTFEGHLMVHPQNY SGKNIPLYISLMGEGSKISNKNFGLITLLMKKNSRPFPSNKKYKIDENVRNMI QPFLSIHFQNTKTYSPYEDSNTMLFPLVLFLTFVGLGIPFNQDGLSKWLT PSINLLHKNSNLLIDWYEFSDKAPFVSIAFGIFIAFLVKPVYSSFNQDLFLNAFY KMGPNRIYDKIKNAIYDMSYNRGYDAFYGFETAGIRRLADAFHFFDRRIIDAIPN GVGLSFFVAEVIKSVGGGRISSEYLFVFSVAIFLLIYFLNL"  
BASE COUNT 296 a 130 c 153 g 384 t  
ORIGIN  
Query Match 2.4%; Score 20; DB 8; Length 963;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 616 ATTAATAATGAAGAGAAATT 635  
|||||  
DB 225 ATTAATAATGAAGAGAAATT 244  
RESULT 9  
LOCUS U23181/c 26758 bp DNA linear INV 29-MAY-2002  
DEFINITION. Caenorhabditis elegans cosmid ZK84, complete sequence.  
ACCESSION U23181  
VERSION U23181.1 GI:726435  
KEYWORDS HTG.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
REFERENCE 1 (bases 1 to 26758)  
AUTHORS Waterston,R.  
TITLE Genome sequence of the nematode C. elegans: a platform for

JOURNAL MEDLINE PUBMED REFERENCE AUTHORS JOURNAL REFERENCE AUTHORS TITLE JOURNAL	investigating biology. The C. elegans Sequencing Consortium Science 282 (5396), 2012-2018 (1998) 99069613 9851916 2 (bases 1 to 26758) Kirsten,J. The sequence of C. elegans cosmid ZK84 Unpublished (2001) 3 (bases 1 to 26758) Waterston,R. Direct Submission Submitted (29-MAY-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA	gene	<pre>/codon_start=1 /product="C. elegans INS-5 protein (corresponding sequence ZK84.3)" /protein_id="AAC48209.1" /db_xref="GI:2914124" /translation="MHSIVALMLIGTILPIAALHOKHQGFTLSSDSTGNQPMDAISR ADPHTNYRSCALGSSRMSENGIDVAOKCCSTDCSSDYIKEICCFD" 4009..4401 /gene="ins-6" /note="for a graphical representation of this gene see: http://www.wormbase.org/db/seq/sequence?name=ZK84.6;class= Sequence" join(4009..4267,4322..4401) /gene="ins-6" /note="Contains similarity to Pfam domain: PF00049 (Insulin), Score=15.3, E-value=0.0041, N=1" /codon_start=1 /product="C. elegans INS-6 protein (corresponding sequence ZK84.6)" /protein_id="AAC48208.1" /db_xref="GI:2914123" /translation="MNSVFTTIIFVLCALQVAASFQSGPSMESASMOLLRELOHN MMESAHPRMPRRARVPAPGETRACGKLISLVMAVCGDLCPQBGKDIATECCGQCS DDYIRSAACP" complement(11974..12278) /gene="ZK84.4" /note="for a graphical representation of this gene see: http://www.wormbase.org/db/seq/sequence?name=ZK84.4;class= Sequence" complement(join(11974..12134,12221..12278)) /gene="ZK84.4" /codon_start=1 /product="Hypothetical protein ZK84.4" /protein_id="AAC48206.1" /db_xref="GI:726439" /translation="MHEQPLEIRFLKFNFPDFPCSKELQELQOKKRTCTTQEFLI LKFLEFLKFTLIKLNLIIPSIPHCCI" 18023..19333 /gene="ZK84.2" /note="for a graphical representation of this gene see: http://www.wormbase.org/db/seq/sequence?name=ZK84.2;class= Sequence" join(18023..18170,18541..18831,19182..19333) /gene="ZK84.2" /note="coded for by the following C. elegans CDNAS: YK207el.5" /codon_start=1 /product="Hypothetical protein ZK84.2" /protein_id="AAC48205.1" /db_xref="GI:726437" /translation="MTEEDVLPKWVLMVLSGMICTYDVAYTMNRYTNDPASPKAN LLFGWALYSSVDHYLTNDVFTGCGRMLECLINFAIGLAFAHSHGLLLAFT SNVWLVKWLILLESFVLQPEGLHPPMNPKNFVWSKEMIFWIPGVMVLPFLVLFAL WNKLALFPKMSKIKWASNKRCGLDSFGDQKRIALE" complement(19772..20026) /gene="ZK84.5" /note="for a graphical representation of this gene see: http://www.wormbase.org/db/seq/sequence?name=ZK84.5;class= Sequence" complement(19772..20026) /gene="ZK84.5" /codon_start=1 /product="Hypothetical protein ZK84.5" /protein_id="AAC48207.1" /db_xref="GI:726440" /translation="MSOMQTLNSKSIKRGPNPKVWIQDNGHEKKTAPAAPTPKTEVT APPTEGFKIKDEVEMLRAGFKNHGKAVKSGQGTMIEL" 24170..26737 /gene="ZK84.1" /note="for a graphical representation of this gene see: http://www.wormbase.org/db/seq/sequence?name=ZK84.1;class= Sequence" join(24170..24234,24288..24357,24411..24597,24654..26737)</pre>
COMMENT	<p>Genome Sequencing Center Department of Genetics, Washington University St. Louis, MO 63110, USA, and Sanger Centre, Hinxton Hall Cambridge CB10 1RQ, England email: rwnematode.wustl.edu and jes@sanger.ac.uk</p> <p>NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.</p> <p>This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality &gt;= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.</p> <p>For a graphical representation of this cosmid sequence and its analysis see: http://www.wormbase.org/db/seq/sequence?name=ZK84;class=Sequence</p> <p>NEIGHBORING COSMID INFORMATION</p> <p>The 5' cosmid is ZK75, 400 bp overlap; the 3' cosmid is R12C12, 2000 bp overlap. Actual start of this cosmid is at base position 1 of ZK84; actual end is at 12202 of R12C12.</p> <p>NOTES:</p> <p>Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFeome cloning project (http://wormfdb.dfc.harvard.edu/), similarity to other proteins from BlastX analyses (http://blast.wustl.edu/), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucleic Acids. Res., 25, 955-964).</p> <p>Location/Qualifiers</p> <p>1..26758 /organism="Caenorhabditis elegans" /strain="Bristol N2" /db_xref="taxon:6239" /chromosome="II" /clone="ZK84" 2002..2352 /gene="ins-5" /note="for a graphical representation of this gene see: http://www.wormbase.org/db/seq/sequence?name=ZK84.3;class= Sequence" join(2002..2188,2267..2352) /gene="ins-5"</p>	gene	
CDS			



/gene="zk84.1"  
/note="final exon in repeat region; similar to long tandem repeat region of sialidase (SP:TCNA\_TRYCR, P23253) and neurofilament H protein; coded for by the following C. elegans cDNAs: yk56a10.3, yk56a10.5, CESE54f"  
/codon\_start=1  
/product="Hypotheetical protein zk84.1"  
/protein\_id="AAC48204.1"  
/db\_xref="GI:726436"  
/translation="MRFGLAVLVVAVYAAPLPVEIKRLDFLKRKTIGEVAIN DTGTGPIKIDNDNLDIFKPRPEIKLRQERDNNAAAASKTDAAVNSASVDYGYG DGAVADVAPGPEPTAPTAPEGAAPVPTAVEGSGVSGIPDEVAITVAADSGSGD APASNSVETLITATEPALIPPAEPKRIIPPSGIVTSQEQQAGADAPAPVVEE TPAPPAAEETAPATEASNAVGTPEGVYDGTANAASVAPEAVETPAPAAEETPA PATSEETAPAPAEETPAPETVSAPEANYSAGDAATAPAPSEADAAPAT DSAASADTAAALVDTSSEHAAEETAPATAETTPAPSPAPVAPVADAAAGYDPS SIPEETPAPAEETPAPASAAEETPAPAAEETPAPETASAPDAAGAAPADVAA PADVATAPETSSAAGSYDVPSEASEVTAPIVESATEAPSDSAIPGPAASEPA PAPIEAPATDAATLETAPAPAEAPAPAAAGYDAPSPVEETPAPAPAADETAPA PAEETPAPAPAEETPAPAPAEETPAPAPAEETPAPAPAEETPAPAPAEETPA PAPAVEETPAPAPAEETPAPAPAEETPAPAPAAEYAAPVAEETPAPAPAEETPA PQAEEETPAPAPAEETPAPAPAEETPAPAPAAEAPAPAPAPETPAPAAEGAAA PAPIVSSGYDAQSTDTVVASSAPATSSDASGATNVEGQQAIFQIARKKVRRAFFGVG GRA"

BASE COUNT 8374 a 5213 c 4710 g 8461 t  
ORIGIN

Query Match 2.4%; Score 20; DB 3; Length 26758;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 622 AATGAAGAAGAAATTGAGAA 641  
|||||  
Db 22320 AATGAAGAAGAAATTGAGAA 22301

RESULT 10  
AL137004  
LOCUS  
DEFINITION  
Human DNA sequence from clone Rpl-216A11 on chromosome 6. Contains ESTs and GSSs, complete sequence.  
ACCESSION  
AL137004  
VERSION  
AL137004.7  
KEYWORDS  
HTG.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 52994)  
Tracey,A.  
Direct Submission  
Submitted (14-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Jul 28, 2000 this sequence version replaced gi:9367914.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information on the WORMPEP database can be found at  
http://www.sanger.ac.uk/projects/Celegans/wormpep  
This sequence was generated from part of bacterial clone contigs of human

chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6  
Rpl-216A11 is from the library RPCI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>  
VECTOR: pCYPAC2  
IMPORTANT: This sequence is not the entire insert of clone Rpl-216A11 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
The true left end of clone Rpl-216A11 is at 1 in this sequence. The true left end of clone Rpl-135L22 is at 52895 in this sequence.

FEATURES  
                    Location/Qualifiers  
                    1..52994  
                    /organism="Homo sapiens"  
                    /db\_xref="taxon:9606"  
                    /chromosome="6"  
                    /clone="Rpl-216A11"  
                    /clone\_lib="RPCI-1"  
                    2076..2303  
                    /note="MIR repeat: matches 8..242 of consensus"  
                    2570..3386  
                    /note="L1PAL6 repeat: matches 5332..6157 of consensus"  
                    complement(3036..3831)  
                    /note="match: GSS: Em:AQ895852"  
                    3511..3824  
                    /note="AluY repeat: matches 1..310 of consensus"  
                    4305..4444  
                    /note="L1MB3 repeat: matches 5685..5828 of consensus"  
                    4481..4790  
                    /note="L1MB2 repeat: matches 5817..6125 of consensus"  
                    4892..4963  
                    /note="AluJ/FRAM repeat: matches 221..292 of consensus"  
                    5109..5389  
                    /note="AluX repeat: matches 18..299 of consensus"  
                    complement(5256..5528)  
                    6277..6587  
                    /note="AluY repeat: matches 1..310 of consensus"  
                    6679..7103  
                    /note="MLTIC repeat: matches 22..464 of consensus"  
                    8056..8313  
                    /note="match: GSS: Em:AQ614111"  
                    8547..8839  
                    /note="AluJo repeat: matches 1..296 of consensus"  
                    9425..9780  
                    /note="L2 repeat: matches 1420..1815 of consensus"  
                    9881..9953  
                    /note="L2 repeat: matches 1686..1750 of consensus"  
                    10202..10226  
                    /note="MLT1J repeat: matches 111..135 of consensus"  
                    10250..10468  
                    /note="MER58A repeat: matches 1..222 of consensus"  
                    10683..10718  
                    /note="18 copies 2 mer tg 86% conserved"  
                    10719..11012  
                    /note="AluJo repeat: matches 1..285 of consensus"  
                    11034..11342  
                    /note="AluX repeat: matches 1..310 of consensus"  
                    11753..12242  
                    /note="match: GSS: Em:AQ604470"  
                    12636..12944  
                    /note="AluX repeat: matches 1..312 of consensus"  
                    complement(13241..13718)  
                    /note="match: GSS: Em:AQ453999"  
                    13488..13792  
                    /note="AluX repeat: matches 1..305 of consensus"  
                    13810..13979  
                    /note="L1MC3 repeat: matches 7546..7739 of consensus"  
                    14860..14935  
                    /note="18 copies 2 mer tt 65% conserved"  
                    15552..15716  
                    /note="L1MB3 repeat: matches 6021..6184 of consensus"

repeat_region	/note="L1MB7 repeat: matches 2983. .3475 of consensus"
repeat_region	39291. .39612
repeat_region	/note="MER46B repeat: matches 1. .236 of consensus"
repeat_region	39613. .41393
repeat_region	/note="L1MB7 repeat: matches 3475. .5249 of consensus"
repeat_region	41400. .41429
repeat_region	/note="15 copies 2 mer ct 93% conserved"
repeat_region	41453. .42323
repeat_region	/note="L1MB7 repeat: matches 5224. .6173 of consensus"
repeat_region	42335. .42645
repeat_region	/note="ALUSx repeat: matches 1. .308 of consensus"
repeat_region	43619. .43872
repeat_region	/note="ALUSg repeat: matches 23. .274 of consensus"
repeat_region	44087. .44381
repeat_region	/note="AluJg repeat: matches 1. .293 of consensus"
repeat_region	44743. .45073
repeat_region	/note="L1PAL1 repeat: matches 5825. .6157 of consensus"
repeat_region	45074. .45254
repeat_region	/note="L1P4 repeat: matches 5610. .5790 of consensus"

  

Query Match	2.4%	Score 20;	DB 9;	Length 52994;
Best Local Similarity	100.08;	Pred. No. 13;		
Matches	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

  

QY	607	TTTTCAGGGATTAAAAATGA	626
Db	20884	TTTTCAGGGATTAAAAATGA	20903

  

RESULT 11	AC098131/c
LOCUS	AC098131
DEFINITION	Rattus norvegicus clone CH230-15517, *** SEQUENCING IN PROGRESS
ACCESSION	AC098131
VERSION	AC098131.4 GI:21729185
KEYWORDS	HTG; HTGS; PHASE1.
SOURCE	Norway rat.
ORGANISM	Rattus norvegicus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
AUTHORS	1 (bases 1 to 95315) Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayelle,M., Banks,T., Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouay,C., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., DeLanaye,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C., Elhaj,C., Escotto,M., Fallis,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabis,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jella,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,L., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,N., Nelson,S., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y.

Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,  
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 95315)  
 Worley, K.C.  
 Direct Submission  
 Submitted (23-OCT-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 95315)  
 Worley, K.C.  
 Direct Submission  
 Submitted (12-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 11, 2002 this sequence version replaced gi:17975697.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GGMA  
 Center clone name: CH230-155I7  
 ----- Summary Statistics  
 ----- Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 24703 bases at least Q40  
 Consensus quality: 27404 bases at least Q30  
 Consensus quality: 29760 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 56 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 1228: contig of 1228 bp in length  
 \* 1229 1328: gap of unknown length  
 \* 1329 2380: contig of 1052 bp in length  
 \* 2381 2480: gap of unknown length  
 \* 2481 4025: contig of 1545 bp in length  
 \* 4026 4125: gap of unknown length  
 \* 4126 5273: contig of 1148 bp in length  
 \* 5274 5373: gap of unknown length  
 \* 5374 6792: contig of 1419 bp in length  
 \* 6793 6893: gap of unknown length  
 \* 6893 8263: contig of 1371 bp in length  
 \* 8264 9584: gap of unknown length  
 \* 9585 9684: gap of unknown length  
 \* 9685 11200: contig of 1516 bp in length  
 \* 11201 11300: gap of unknown length  
 \* 11301 12388: contig of 1088 bp in length  
 \* 12389 12489: gap of unknown length  
 \* 12489 13555: contig of 1067 bp in length  
 \* 13556 13656: gap of unknown length  
 \* 13656 14809: contig of 1154 bp in length  
 \* 14810 14909: gap of unknown length

16044: contig of 1135 bp in length  
 16144: gap of unknown length  
 16145 17476: contig of 1332 bp in length  
 17477 17576: gap of unknown length  
 17577 18689: contig of 1113 bp in length  
 18690 18789: gap of unknown length  
 18790 20014: contig of 1225 bp in length  
 20015 20114: gap of unknown length  
 20115 21754: contig of 1640 bp in length  
 21755 21854: gap of unknown length  
 21855 23225: contig of 1371 bp in length  
 23226 23225: gap of unknown length  
 23226 25003: contig of 1678 bp in length  
 25004 25103: gap of unknown length  
 25104 26236: contig of 1133 bp in length  
 26237 26336: gap of unknown length  
 26337 27859: contig of 1523 bp in length  
 27860 27959: gap of unknown length  
 27960 29863: contig of 1904 bp in length  
 29864 31499: contig of 1536 bp in length  
 31500 31599: gap of unknown length  
 31600 33222: contig of 1623 bp in length  
 33223 33222: gap of unknown length  
 33223 34463: contig of 1141 bp in length  
 34464 34563: gap of unknown length  
 34564 35733: contig of 1170 bp in length  
 35734 35833: gap of unknown length  
 35834 37005: contig of 1172 bp in length  
 37006 37105: gap of unknown length  
 37106 38260: contig of 1155 bp in length  
 38261 38360: gap of unknown length  
 38361 39391: contig of 1031 bp in length  
 39392 39491: gap of unknown length  
 39492 40765: contig of 1274 bp in length  
 40766 40865: gap of unknown length  
 40866 42016: contig of 1151 bp in length  
 42017 42116: gap of unknown length  
 42117 43321: contig of 1205 bp in length  
 43322 43421: gap of unknown length  
 43422 43422: contig of 1708 bp in length  
 43423 45129: contig of 1708 bp in length  
 45130 45229: gap of unknown length  
 45230 47027: contig of 1798 bp in length  
 47028 47127: gap of unknown length  
 47128 48225: contig of 1098 bp in length  
 48226 48325: gap of unknown length  
 48326 50301: contig of 1976 bp in length  
 50302 50401: gap of unknown length  
 50402 52152: contig of 1751 bp in length  
 52153 52252: gap of unknown length  
 52253 53970: contig of 1718 bp in length  
 53971 54070: gap of unknown length  
 54071 55268: contig of 1198 bp in length  
 55269 55368: gap of unknown length  
 55369 56556: contig of 1288 bp in length  
 56557 56756: gap of unknown length  
 56757 58254: contig of 1498 bp in length  
 58255 58354: gap of unknown length  
 58355 59445: contig of 1091 bp in length  
 59446 59545: gap of unknown length  
 59546 62045: contig of 2500 bp in length  
 62046 62145: gap of unknown length  
 62146 64566: contig of 2421 bp in length  
 64567 64666: gap of unknown length  
 64667 66958: contig of 2292 bp in length  
 66959 67058: gap of unknown length  
 67059 69063: contig of 2005 bp in length  
 69064 69163: gap of unknown length  
 69164 71523: contig of 2360 bp in length  
 71524 71623: gap of unknown length  
 71624 73080: contig of 1457 bp in length  
 73081 73180: gap of unknown length  
 73181 74950: contig of 1770 bp in length

```

* 74951 75050: gap of unknown length
* 75051 77416: contig of 2366 bp in length
* 77417 77516: gap of unknown length
* 77517 78338: contig of 1422 bp in length
* 78339 79038: gap of unknown length
* 79039 81846: contig of 2808 bp in length
* 81847 81946: gap of unknown length
* 81947 84023: contig of 2077 bp in length
* 84024 84123: gap of unknown length
* 84124 86119: contig of 1996 bp in length
* 86120 86219: gap of unknown length
* 86220 87407: contig of 1188 bp in length

Query Match      2.4%; Score 20; DB 2; Length 95315;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 AAGAAGCTAGAGGAAG 43
|||||
Db 45574 AAGAAGCTAGAGGAAG 45555

RESULT 12
AC104680/c
LOCUS
DEFINITION
Sus scrofa clone RP44-434D6, WORKING DRAFT SEQUENCE, 5 unordered
pieces.
ACCESSION
AC104680.1 GI:17933850
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
Sus scrofa.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
1 (bases 1 to 150877)
AUTHORS
Akhter N., Ayele K., Beckstrom-Sternberg S.M., Benjamin B.,
Blakesley P.W., Bouffard G.G., Breen K., Brinkley C., Brooks S.,
Dietrich N.L., Granite S., Guan X., Gupta J., Haghighi P.,
Ho S.-I., Idoi J.R., Karlins E., Laric P., Lee Lin S.-Q.,
Legasp, R., Maduro G.L., Maduro V.B., Mashello C., Mastrian S.D.,
McCluskey J.C., McDowell J., Pearson R., Prasad A., Stantripo S.,
Thomas J.W., Thomas P.J., Touchman J.W., Tsurgeon C., Vogt J.L.,
Walker M.A., Wetherby K.D., Wiggins L., Young A., Zhang L.-H. and
Green E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 150877)
Green, E.D.
Direct Submission
Submitted (19-DEC-2001) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@hghri.nih.gov
----- Project Information
Center project name: Cnn
Center clone name: 434D6
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 148948 bases at least Q40
Consensus quality: 149609 bases at least Q30
Consensus quality: 150008 bases at least Q20
Insert size: 147000; agarose-fp
Insert size: 150477; sum-of-contigs
Quality coverage: 12.81x in Q20 bases; agarose-fp
Quality coverage: 12.51x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces

```

```

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 4912: contig of 4912 bp in length
* 4913 5012: gap of unknown length
* 5013 30610: contig of 25598 bp in length
* 30611 30710: gap of unknown length
* 30711 49660: contig of 18950 bp in length
* 49661 49760: gap of unknown length
* 49761 86175: contig of 36415 bp in length
* 86176 86275: gap of unknown length
* 86276 150877: contig of 64602 bp in length.

FEATURES
source
1..150877
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="RP44-434D6"
/clone_lib="RP44"
1..4912
/note="assembly_fragment"
5013..30610
/note="assembly_fragment"
clone_end:77
vector_side:right
30711..49660
/note="assembly_fragment"
clone_end:SP6
vector_side:right
49761..86175
/note="assembly_fragment"
86276..150877
/note="assembly_fragment"
BASE COUNT 45369 a 29612 c 30757 g 44733 t 406 others
ORIGIN
Query Match      2.4%; Score 20; DB 2; Length 150877;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 784 AATCTGTAAATCAAGTGT 803
|||||
Db 130771 AATCTGTAAATCAAGTGT 130752

RESULT 13
AC021284
LOCUS
DEFINITION
Homo sapiens clone RP11-21G24, WORKING DRAFT SEQUENCE, 18 unordered
pieces.
ACCESSION
AC021284.3 GI:7249031
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
Homo sapiens.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 176699)
AUTHORS
Birren B., Linton L., Nusbaum C. and Lander E.
TITLE
Homo sapiens, clone RP11-21G24
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 176699)
AUTHORS
Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,
Anderson S., Baldwin J., Barna N., Beckerly R., Beda F.,
Boguslavsky L., Boukhgalter B., Brown A., Burkett G., Castle A.,
Choepel Y., Collangelo M., Collins S., Collymore A., Cooke P.,
DeArelano K., Dewar K., Domino M., Doyle M., Fenesfor J.,
Ferreira P., FitzHugh W., Forrest C., Gage D., Galagan J.,
Gardyna S., Grant G., Hagos B., Heaford A., Horton L.,
Howland J.C., Johnson R., Jones C., Kann L., Karatas A., Klein J.,
Landers T., Lehoczy J., Levine R., Lieu C., Liu G., Locke K.,

```

**TITLE**  
**JOURNAL**  
**COMMENT**

Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,  
McPheeters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,  
Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, T.M., Peterson, K.,  
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severi, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,  
Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 16, 2000 this sequence version replaced gi:5970503.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L4010  
Center clone name: 21.G-24  
----- Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 164878 bases at least Q40  
Consensus quality: 170409 bases at least Q30  
Consensus quality: 172980 bases at least Q20  
Insert size: 180000; agarose-fp  
Insert size: 174999; sum-of-contigs  
Quality coverage: 4.4 in Q20 bases; agarose-fp  
Quality coverage: 4.5 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 18 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 1319: contig of 1319 bp in length  
\* 1320 1419: gap of 100 bp  
\* 1420 4237: contig of 2808 bp in length  
\* 4238 4327: gap of 100 bp  
\* 4328 7630: contig of 3303 bp in length  
\* 7631 7730: gap of 100 bp  
\* 7731 10250: contig of 2520 bp in length  
\* 10251 10350: gap of 100 bp  
\* 10351 13714: contig of 3364 bp in length  
\* 13715 13814: gap of 100 bp  
\* 13815 16659: contig of 2845 bp in length  
\* 16660 16759: gap of 100 bp  
\* 16760 21588: contig of 4829 bp in length  
\* 21589 21688: gap of 100 bp  
\* 21689 25153: contig of 3465 bp in length  
\* 25154 25253: gap of 100 bp  
\* 25254 29882: contig of 4629 bp in length  
\* 29883 29982: gap of 100 bp  
\* 29983 34176: contig of 4194 bp in length  
\* 34177 34276: gap of 100 bp  
\* 34277 39604: contig of 5328 bp in length  
\* 39605 39704: gap of 100 bp  
\* 39705 48245: contig of 8541 bp in length  
\* 48246 48345: gap of 100 bp  
\* 48346 57261: contig of 8916 bp in length  
\* 57262 57361: gap of 100 bp  
\* 57362 74202: contig of 16841 bp in length  
\* 74203 74302: gap of 100 bp  
\* 74303 92606: contig of 18304 bp in length  
\* 92607 92706: gap of 100 bp

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 180748)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus chromosome, clone RP23-343H21
Unpublished
2 (bases 1 to 180748)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Canarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Tsch,K.,
Li,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 180748)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Canarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Tsch,K.,
Li,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 16, 2002 this sequence version replaced gi:22004418.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23770
Center clone name: 343_H_21
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 179045 bases at least Q40
Consensus quality: 179810 bases at least Q30
Consensus quality: 180026 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 180148; sum-of-contigs
Quality coverage: 8.5 in Q20 bases; agarose-fp
Quality coverage: 8.6 in Q20 bases; sum-of-contigs

```

```

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 13353: contig of 13353 bp in length
* 13354 13453: gap of 100 bp
* 13454 15102: contig of 1649 bp in length
* 15103 15202: gap of 100 bp
* 15203 20389: contig of 5187 bp in length
* 20390 20489: gap of 100 bp
* 20490 34118: contig of 13629 bp in length
* 34119 34218: gap of 100 bp
* 34219 52624: contig of 18406 bp in length
* 52625 52724: gap of 100 bp
* 52725 137521: contig of 84797 bp in length
* 137522 137621: gap of 100 bp
* 137622 180748: contig of 43127 bp in length.
*
* Location/Qualifiers
* 1..180748
* /organism="Mus musculus"
* /db_xref="taxon:10090"
* /clone="RP23-343H21"
* /clone_lib="RPCI-23 Female Mouse BAC"
* 1..13353
* /note="assembly_fragment"
* clone_end:SP6
* vector_side:left"
* 13454..15102
* /note="assembly_fragment"
* 15203..20389
* /note="assembly_fragment"
* 20490..34118
* /note="assembly_fragment"
* 34219..52624
* /note="assembly_fragment"
* 52725..137521
* /note="assembly_fragment"
* 137622..180748
* /note="assembly_fragment"
* clone_end:T7
* vector_side:right"
*
* BASE COUNT 49564 a 38649 c 39535 g 52399 t 601 others
* ORIGIN
*
* Query Match 2.4%; Score 20; DB 2: Length 180748;
* Best Local Similarity 100.0%; Pred. No. 12;
* Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* QY 24 AAGAAGGCTAGAGGAAAG 43
* |||||
* Db 167531 AAGAAGGCTAGAGGAAAG 167550
*
* RESULT 15
* AC011742
* LOCUS Homo sapiens BAC clone RF11-214J9 from 2, complete sequence.
* DEFINITION AC011742
* ACCESSION AC011742
* VERSION AC011742.3 GI:9887800
* KEYWORDS HTG.
* SOURCE Homo sapiens.
* ORGANISM Homo sapiens
* Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
* Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
* 1 (bases 1 to 181210)
* Sulston,J.E. and Waterston,R.
* Toward a complete human genome sequence
* Genom Res. 8 (11), 1097-1108 (1998)
* JOURNAL

```



```
repeat_region 9329. .9529
/rpt_family="L1"
repeat_region 9525. .9734
/rpt_family="L1"
repeat_region 9980. .10069
/rpt_family="MIR"
repeat_region 10070. .10375
/rpt_family="Alu"
repeat_region 10376. .10546
/rpt_family="MIR"
misc_feature 10882. .10914
note="similar to EST AAL48206 (NID:gl717613) zo53f11.sl"
repeat_region 10883. .11204
/rpt_family="Alu"
repeat_region 11209. .11479
/rpt_family="Alu"
repeat_region 11556. .11680
/rpt_family="Alu"
repeat_region 12186. .12495
/rpt_family="Alu"
repeat_region 12622. .12921
/rpt_family="Alu"
repeat_region 13091. .13564
/rpt_family="L1"
repeat_region 13593. .13844
/rpt_family="Alu"
repeat_region 13851. .13993
/rpt_family="L1"
repeat_region 14123. .14435
/rpt_family="Alu"
repeat_region 14554. .14850
/rpt_family="Alu"
repeat_region 15228. .15339
/rpt_family="ERV"
misc_feature 15397. .15989
note="similar to EST BE080155 (NID:g8470442)"
repeat_region 16224. .16562
/rpt_family="MALR"
repeat_region 16567. .16865
/rpt_family="Alu"
repeat_region 16869. .17004
```

```
Query Match 2.4% Score 20; DB 9; Length 181210;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 TCAACCTATGGAAACTGGA 269
Db 87347 TCAACCTATGGAAACTGGA 87366
|||||
```

Search completed: February 18, 2003, 19:33:48  
Job time : 3385 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 02:02:38 ; Search time 251 Seconds  
(without alignments)  
7348.154 Million cell updates/sec

Title: US-09-673-918A-1

Perfect score: 819  
Sequence: 1 atgcagcttcgaactgattt.....tggccaataatccagattct 819

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	819	100.0	819	21	AAZ45730
2	816	99.6	816	21	AAZ45743
3	806.4	98.5	831	21	AAZ45731
4	736	89.9	828	21	AAZ45734
5	736	89.9	4505	24	AAZ38432
6	733.8	89.6	819	21	AAZ45732
7	712	86.9	831	21	AAZ45733
8	200.2	24.4	1120	21	AAZ40811
9	188.8	23.1	1025	16	AAQ94623

10	187	22.8	1136	21	AAZ38191	Arabidopsis thalia
11	185.4	22.6	852	21	AAZ43068	Arabidopsis thalia
12	183.6	22.4	1138	21	AAZ45620	Arabidopsis thalia
13	150.2	18.3	774	21	AAZ43159	Arabidopsis thalia
14	149	18.2	833	24	ABN98918	Arabidopsis thalia
15	148.4	18.1	774	21	AAZ50609	Arabidopsis thalia
16	147.6	18.0	1002	21	AAZ43363	Arabidopsis thalia
17	146.2	17.9	997	21	AAZ50610	Arabidopsis thalia
18	145.4	17.8	861	21	AAZ50608	Arabidopsis thalia
19	99.8	12.2	941	21	AAZ34703	Arabidopsis thalia
20	98.2	12.0	863	21	AAZ98308	A. thaliana gene i
21	97.2	11.9	1236	16	AAZ45060	Maize Ts2 cDNA nuc
22	90	11.0	338	21	AAZ57005	Pinus radiata tran
23	88.4	10.8	421	21	AAZ56947	Pinus radiata tran
24	88.4	10.8	497	21	AAZ56998	Pinus radiata tran
25	86.8	10.6	440	21	AAZ56990	Pinus radiata tran
26	86.8	10.6	453	21	AAZ56982	Pinus radiata tran
27	86.8	10.6	475	21	AAZ57080	Pinus radiata tran
28	86.2	10.5	441	21	AAZ57063	Pinus radiata tran
29	85.8	10.5	323	21	AAZ57065	Pinus radiata tran
30	85.6	10.5	465	22	AAH87673	Peppermint plant o
31	82.6	10.1	498	21	AAZ57104	Pinus radiata tran
32	80	9.8	1076	24	ABL56024	Plodia interpuncte
33	79.6	9.7	903	24	ABQ69008	Listeria monocytog
34	79.6	9.7	8307	24	ABQ71075	Listeria monocytog
35	78	9.5	6948	20	AAZ12941	Enterococcus faeca
36	73.2	8.9	810	24	ABN67211	Streptococcus poly
37	71.8	8.8	311	21	AAZ56919	Pinus radiata tran
38	71.8	8.8	319630	24	ABQ67194	Listeria innocua c
39	68.6	8.4	221	21	AAZ57012	Pinus radiata tran
40	67.4	8.2	7471	21	AAZ88789	B. subtilis bac DN
41	67	8.2	221	21	AAZ56958	Pinus radiata tran
42	65.2	8.0	4281	24	ABA91715	Comamonas sp. cycl
43	62.6	7.6	306	21	AAZ56932	Pinus radiata tran
44	62.6	7.6	2944528	24	ABA03041	Listeria monocytog
45	62.4	7.6	774	24	AAD28464	Bacillus subtilis

ALIGNMENTS

RESULT 1	
AAZ45730	
ID	AAZ45730 standard; cDNA; 819 BP.
XX	
AC	AAZ45730;
XX	
DT	06-APR-2000 (first entry)
XX	
DE	Secoisolaricresinol dehydrogenase cDNA clone DEHY133.
XX	
KW	Secoisolaricresinol dehydrogenase; lignan biosynthetic pathway;
KW	lignan; metairesinol; health-protecting lignan; phytoestrogen;
KW	enterolactone; enterodiol; neutriceutical; dietary supplement;
KW	(-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.
XX	
OS	Forsythia intermedia.
XX	
Key	Location/Qualifiers
FT	1..819
CDS	
FT	/tag= a
FT	/product= "secoisolaricresinol dehydrogenase"
FT	/note= "no termination codon given"
XX	
XX	WO9955846-A1.
PN	
XX	
XX	04-NOV-1999.
PD	
XX	
XX	23-APR-1999; 99WO-US08975.
PF	
XX	
XX	24-APR-1998; 98US-0082977.
PR	
XX	
XX	(UNIW ) UNIV WASHINGTON STATE RES FOUND.

XX Xia Z, Costa MA, Davin LB, Lewis NG;  
 XX WPI: 2000-126356/11.  
 DR P-PSDB: AAY54412.  
 XX  
 PT New nucleic acid molecule encoding an enzyme involved in lignan  
 PT biosynthetic pathway, useful for producing large amounts of lignans  
 XX  
 PS Claim 9: Page 45-46; 66pp; English.  
 XX  
 CC The present sequence encodes a secoisolariciresinol dehydrogenase  
 CC protein. The enzyme is involved in the lignan biosynthetic pathway.  
 CC The secoisolariciresinol dehydrogenase proteins have a molecular  
 CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require  
 CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase  
 CC nucleic acids are used for the recombinant expression of the enzymes.  
 CC It is also used to obtain expression or enhanced expression of  
 CC secoisolariciresinol dehydrogenase in plants or to alter lignan  
 CC biosynthesis. The enzyme is used for production of the pharmacologically  
 CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase  
 CC proteins and nucleic acids can be utilized to: elevate or otherwise  
 CC alter the levels of health-protecting lignans, including phytoestrogens  
 CC such as enterolactone and enterodiol, in plant species, including  
 CC vegetables, grains and fruits and to food items incorporating material  
 CC derived from such genetically altered plants; genetically alter plant  
 CC species to provide an abundant, natural supply of lignans useful for  
 CC a variety of purposes, for example as nutraceuticals and dietary  
 CC supplements; to genetically alter living organisms to produce an  
 CC abundant supply of optically pure lignans having desirable biological  
 CC properties, for example, (-)-trachelogenin which possesses antiviral  
 CC properties, and (-)-podophyllotoxin.  
 XX  
 SQ Sequence 819 BP; 240 A; 166 C; 189 G; 224 T; 0 other;  
 Query Match 100.0%; Score 819; DB 21; Length 819;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-230;  
 Matches 819; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTTCGAACTCGCAAGAGGCTAGAGGAAAAGTTGCCCTTATACAGGA 60  
 DB 1 ATGCAGCTTCGAACTCGCAAGAGGCTAGAGGAAAAGTTGCCCTTATACAGGA 60  
 QY 61 GGAGCCAGTGAATTTGGAGAAACACAGCAAAACTCTTCCCAACATGGAGCCAAAGTT 120  
 DB 61 GGAGCCAGTGAATTTGGAGAAACACAGCAAAACTCTTCCCAACATGGAGCCAAAGTT 120  
 QY 121 GCCATTGCTGATGTCGCAAGATGAATAGTCACTCAGTTGTCGAGGCCATTGGCAGCTCC 180  
 DB 121 GCCATTGCTGATGTCGCAAGATGAATAGTCACTCAGTTGTCGAGGCCATTGGCAGCTCC 180  
 QY 181 AATTCACCTACATCCACTGTGATGTTACTAATGAAGAGCGGTGTTAAATAATGCCGTGGAC 240  
 DB 181 AATTCACCTACATCCACTGTGATGTTACTAATGAAGAGCGGTGTTAAATAATGCCGTGGAC 240  
 QY 241 AACACAGTTCAACCTATGAAACACTGGACATATATGTCAGCAATGCAGGAATTTCTGAT 300  
 DB 241 AACACAGTTCAACCTATGAAACACTGGACATATATGTCAGCAATGCAGGAATTTCTGAT 300  
 QY 301 CCCAACAGGCCGCCCATCATAGACAAACGAAAGAGAGACTTTGAACGCGTCTTCAGTGTA 360  
 DB 301 CCCAACAGGCCGCCCATCATAGACAAACGAAAGAGAGACTTTGAACGCGTCTTCAGTGTA 360  
 QY 361 AATGTAACCGGAGTTTTCCTATCATGAGCAGCAGCGGTGTTATCATTCAGCAGCGC 420  
 DB 361 AATGTAACCGGAGTTTTCCTATCATGAGCAGCAGCGGTGTTATCATTCAGCAGCGC 420  
 QY 421 AGTGGACACATATTCCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCT 480  
 DB 421 AGTGGACACATATTCCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCT 480  
 QY 481 GCCTATTGTTGGTTCAAGAGATGCTGTGTAGCCCTTACTAGGAATCTCGCAGTCCAGCTC 540  
 DB 481 GCCTATTGTTGGTTCAAGAGATGCTGTGTAGCCCTTACTAGGAATCTCGCAGTCCAGCTC 540

DB 481 GCCTATTGTTGGTTCAAGAGATGCTGTGTAGCCCTTACTAGGAATCTGGCAGTCCAGCTC 540  
 QY 541 GGACAAATTTGGCATTAGGTTAAATGTTCTCTCCCTTCGGGCTTCTACGGCTTTAGGC 600  
 DB 541 GGACAAATTTGGCATTAGGTTAAATGTTCTCTCCCTTCGGGCTTCTACGGCTTTAGGC 600  
 QY 601 AGAATATTTTCAGGAGTAAATAATGAAGAAGATTTGAGAACTATATAAATTTCCGGGA 660  
 DB 601 AGAATATTTTCAGGAGTAAATAATGAAGAAGATTTGAGAACTATATAAATTTCCGGGA 660  
 QY 661 AATTTGAAGTCCAAAATTTAATGTTGAGGATTTGCCAATGCCAGCTCTTTATCTGGCT 720  
 DB 661 AATTTGAAGTCCAAAATTTAATGTTGAGGATTTGCCAATGCCAGCTCTTTATCTGGCT 720  
 QY 721 AGTATGATGAGCAAAATACGTGAGTGGACAACTCTGTTCAATGATGAGGGTTTCAGCGTC 780  
 DB 721 AGTATGATGAGCAAAATACGTGAGTGGACAACTCTGTTCAATGATGAGGGTTTCAGCGTC 780  
 QY 781 TGCATTTCTGTAATCAAAAGTGTTCCTCAATATCCAGATTCT 819  
 DB 781 TGCATTTCTGTAATCAAAAGTGTTCCTCAATATCCAGATTCT 819

RESULT 2  
 AA245743  
 ID AA245743 standard; CDNA; 816 BP.  
 AC AA245743;  
 XX  
 DT 06-APR-2000 (first entry)  
 XX  
 DE Secoisolariciresinol dehydrogenase cDNA clone DEHY130.  
 XX  
 KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;  
 KW lignan; matairesinol; health-protecting lignan; phytoestrogen;  
 KW enterolactone; enterodiol; nutraceutical; dietary supplement;  
 KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.  
 OS Forsythia intermedia.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..816  
 FT /\*Tag= a  
 FT /product= "secoisolariciresinol dehydrogenase"  
 FT /note= "no termination codon given; partial  
 FT sequence"  
 XX  
 PN W09955846-A1.  
 XX  
 PD 04-NOV-1999.  
 XX  
 PF 23-APR-1999; 99WO-US08975.  
 PR 24-APR-1998; 98US-0082977.  
 XX  
 PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.  
 XX  
 PI Xia Z, Costa MA, Davin LB, Lewis NG;  
 XX  
 DR WPI: 2000-126356/11.  
 DR P-PSDB: AAY54420.  
 XX  
 PT New nucleic acid molecule encoding an enzyme involved in lignan  
 PT biosynthetic pathway, useful for producing large amounts of lignans  
 XX  
 PS Example 2: Page 61-62; 66pp; English.  
 CC The present sequence encodes a secoisolariciresinol dehydrogenase  
 CC protein. The enzyme is involved in the lignan biosynthetic pathway.  
 CC The secoisolariciresinol dehydrogenase proteins have a molecular  
 CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require  
 CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase  
 CC nucleic acids are used for the recombinant expression of the enzymes.

It is also used to obtain expression or enhanced expression of secoisolariciresinol dehydrogenase in plants or to alter lignan biosynthesis. The enzyme is used for production of the pharmacologically active lignan, matairesinol. The secoisolariciresinol dehydrogenase proteins and nucleic acids can be utilized to: elevate or otherwise alter the levels of health-protecting lignans, including phytoestrogens such as enterolactone and enterodiol, in plant species, including vegetables, grains and fruits and to food items incorporating material derived from such genetically altered plants; genetically alter plant species to provide an abundant, natural supply of lignans useful for a variety of purposes, for example as nutraceuticals and dietary supplements; to genetically alter living organisms to produce an abundant supply of optically pure lignans having desirable biological properties, for example (-)-trachelogenin which possesses antiviral properties, and (-)-podophyllotoxin.

Sequence 816 BP; 239 A; 166 C; 188 G; 223 T; 0 other;

Query Match 99.8%; Score 816; DB 21; Length 816;  
Best Local Similarity 100.0%; Pred. No. 2:8e-229;  
Matches 816; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CAGCTCGAACTGCATTCGCAAGAGCTAGAGGAAAGTTGCCCTTATACAGGAGGA 63  
DB 1 CAGCTTCGAAGCTGATTCGCAAGAGGCTAGAGGAAAGTTGCCCTTATACAGGAGGA 60  
QY 64 GCCAGTGGAAATTGGAGAAACACAGCAAAACTCTCTCCCAACATGGAGCAAAAGTTGCC 123  
DB 61 GCCAGTGGAAATTGGAGAAACACAGCAAAACTCTCTCCCAACATGGAGCAAAAGTTGCC 120  
QY 124 ATTGCTGATGTCGAAGATGAATTAGTCTACTCAGTGTGTCGAGGCCATTTGGCACTTCCAAT 183  
DB 121 ATTGCTGATGTCGAAGATGAATTAGTCTACTCAGTGTGTCGAGGCCATTTGGCACTTCCAAT 180  
QY 184 TCCACCTACATCCACTGTGATGTTACTAATGAAGACGGTCTTAAATAATGCCGTGGCAAC 243  
DB 181 TCCACCTACATCCACTGTGATGTTACTAATGAAGACGGTCTTAAATAATGCCGTGGCAAC 240  
QY 244 ACAGTTTCAACCTATGAAACATGGACATTTATGTTGAGCAATGCAGGAATTTCTGATCCC 303  
DB 241 ACAGTTTCAACCTATGAAACATGGACATTTATGTTGAGCAATGCAGGAATTTCTGATCCC 300  
QY 304 AACAGGCCCGCATATAGACACAGCAAAAGACATTTGACACGGCTTCTCAGTGTAAAT 363  
DB 301 AACAGGCCCGCATATAGACACAGCAAAAGACATTTGACACGGCTTCTCAGTGTAAAT 360  
QY 364 GTACCGGAGTTTCCCTATGATGATGACGACGACGCGTGTATGATTCAGCACCCAGT 423  
DB 361 GTACCGGAGTTTCCCTATGATGATGACGACGACGCGTGTATGATTCAGCACCCAGT 420  
QY 424 GGCAACATAATTTCCACTGCTAGTTTAAAGCTCAACTATGGTGGTGGTCTTCCACATGCC 483  
DB 421 GGCAACATAATTTCCACTGCTAGTTTAAAGCTCAACTATGGTGGTGGTCTTCCACATGCC 480  
QY 484 TATGTGGTTCAAAGCATGCTGTGTAGCCCTTACTAGGAATCTGGAGTCGAGCTCGGA 543  
DB 481 TATGTGGTTCAAAGCATGCTGTGTAGCCCTTACTAGGAATCTGGAGTCGAGCTCGGA 540  
QY 544 CAATTTGGCATTAGGTTAATCTGTTCTCTCCTTCGGGCTTCTACGCGCTTTAGGCAAG 603  
DB 541 CAATTTGGCATTAGGTTAATCTGTTCTCTCCTTCGGGCTTCTACGCGCTTTAGGCAAG 600  
QY 604 AAATTTTCAGGGATTAATAATGAAGAGAAATTTGAGAATGTAATAAATTTTCGGGGAAT 663  
DB 601 AAATTTTCAGGGATTAATAATGAAGAGAAATTTGAGAATGTAATAAATTTTCGGGGAAT 660  
QY 664 TTGAAGTCCAAATTTAATGTGAGGATGTTGCCAATGCAGCTTTTATCTGCGGTAGT 723  
DB 661 TTGAAGTCCAAATTTAATGTGAGGATGTTGCCAATGCAGCTTTTATCTGCGGTAGT 720  
QY 724 GATGAGGCAAAATACGTGAGTGACACAAATCTGTTTATTCATGATGAGGGTTACGCGTCTGC 783  
DB 721 GATGAGGCAAAATACGTGAGTGACACAAATCTGTTTATTCATGATGAGGGTTACGCGTCTGC 780

QY 784 AATTCGTGTAATCAAAAGTGTTCCTCAATATCCAGATTCT 819  
DB 781 AATTCGTGTAATCAAAAGTGTTCCTCAATATCCAGATTCT 816

RESULT 3  
AAZ45731  
ID AAZ45731 standard; cDNA; 831 BP.  
XX AAZ45731;  
XX AC AAZ45731;  
DT 06-APR-2000 (first entry)  
XX DE Secoisolariciresinol dehydrogenase cDNA clone SMDEHY321.  
XX KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;  
KW lignan; matairesinol; health-protecting lignan; phytoestrogen;  
KW enterolactone; enterodiol; nutraceutical; dietary supplement;  
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.  
OS Forsythia intermedia.  
XX Key Location/Qualifiers  
FH 1.831  
FT /tag= a  
FT /product= "secoisolariciresinol dehydrogenase"  
FT /note= "no termination codon given"  
XX WO955846-A1.  
XX 04-NOV-1999.  
XX 23-APR-1999; 99WO-US08975.  
XX 24-APR-1998; 98US-0082977.  
XX (UNIW ) UNIV WASHINGTON STATE RES FOUND.  
XX Xia Z, Costa MA, Davin LB, Lewis NG;  
WPI; 2000-126356/11.  
P-PSDB; AAY54413.  
PT New nucleic acid molecule encoding an enzyme involved in lignan biosynthetic pathway, useful for producing large amounts of lignans  
PS Claim 9; Page 48-49; 66pp; English.  
XX The present sequence encodes a secoisolariciresinol dehydrogenase protein. The enzyme is involved in the lignan biosynthetic pathway. The secoisolariciresinol dehydrogenase proteins have a molecular weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase nucleic acids are used for the recombinant expression of the enzymes. It is also used to obtain expression or enhanced expression of secoisolariciresinol dehydrogenase in plants or to alter lignan biosynthesis. The enzyme is used for production of the pharmacologically active lignan, matairesinol. The secoisolariciresinol dehydrogenase proteins and nucleic acids can be utilized to: elevate or otherwise alter the levels of health-protecting lignans, including phytoestrogens such as enterolactone and enterodiol, in plant species, including vegetables, grains and fruits and to food items incorporating material derived from such genetically altered plants; genetically alter plant species to provide an abundant, natural supply of lignans useful for a variety of purposes, for example as nutraceuticals and dietary supplements; to genetically alter living organisms to produce an abundant supply of optically pure lignans having desirable biological properties, for example (-)-trachelogenin which possesses antiviral properties, and (-)-podophyllotoxin.

Sequence 831 BP; 244 A; 169 C; 192 G; 226 T; 0 other;

Query Match 98.5%; Score 806.4; DB 21; Length 831;  
Best Local Similarity 99.3%; Pred. No. 1.9e-226;  
Matches 810; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 CAGCTTCCGAATGCAATGCGAAGAGGCTAGAGGAAAAGTTGCCCTTATAACAGGAGGA 63  
DB 16 CAGGTTCTTAACATGCAATGCGAAGAGGCTAGAGGAAAAGTTGCCCTTATAACAGGAGGA 75

QY 64 GCACGTGAATGGAGAAACACACAGCAAACTCTCTCCACATGGAGCCAAAGTTGCC 123  
DB 76 GCACGTGAATGGAGAAACACACAGCAAACTCTCTCCACATGGAGCCAAAGTTGCC 135

QY 124 ATTGCTGATGCCAAGATGAATAGTCACTAGTGTGCGAGGCCATTGGGCACATTCCTCAAT 183  
DB 136 ATTGCTGATGCCAAGATGAATAGTCACTAGTGTGCGAGGCCATTGGGCACATTCCTCAAT 195

QY 184 TCCACCTACATCCACTGTGATGTTACTAATGAAGACGGTGTAAATAATGCCGTGGACAAC 243  
DB 196 TCCACCTACATCCACTGTGATGTTACTAATGAAGACGGTGTAAATAATGCCGTGGACAAC 255

QY 244 ACAGTTTCAACCTATGGAACACTGGACATTTATGTTCCAGCAATGCAGGAATTTCTGATCCC 303  
DB 256 ACAGTTTCAACCTATGGAACACTGGACATTTATGTTCCAGCAATGCAGGAATTTCTGATCCC 315

QY 304 AACAGCCGCCCATCATAGACAAACGAAAAGCAGACTTTTGAACGGTCTCTCAGTGTAAAT 363  
DB 316 AACAGCCGCCCATCATAGACAAACGAAAAGCAGACTTTTGAACGGTCTCTCAGTGTAAAT 375

QY 364 GTAACCGGAGTTTCCCTATGATGACGACGACGACGTTTATGATCCACGACGCAAT 423  
DB 376 GTAACCGGAGTTTCCCTATGATGACGACGACGACGTTTATGATCCACGACGCAAT 435

QY 424 GGCACATAAATTCACATGCTAGTTTAAGCTCAACTATGCGTGGTGGTCTTCACATGCC 483  
DB 436 GGCACATAAATTCACATGCTAGTTTAAGCTCAACTATGCGTGGTGGTCTTCACATGCC 495

QY 484 TATGCTGTTCAAGCAATGCTGTGTTAGCCCTTACTAGGAATCTGGCAGTCGAGTCGGA 543  
DB 496 TATGCTGTTCAAGCAATGCTGTGTTAGCCCTTACTAGGAATCTGGCAGTCGAGTCGGA 555

QY 544 CAATTTGGCATAGGGTAAATGTTGTCTCCCTTCCGGCTTCTTACGGCTTTTAGGCAAG 603  
DB 556 CAATTTGGCATAGGGTAAATGTTGTCTCCCTTCCGGCTTCTTACGGCTTTTAGGCAAG 615

QY 604 AAATTTCCAGGGATTAATAATGAAGAGAAATTTGAGAATGTATAAATCTTTCGGGAAAT 663  
DB 616 AAATTTCCAGGGATTAATAATGAAGAGAAATTTGAGAATGTATAAATCTTTCGGGAAAT 675

QY 664 TTGAAAGTCCAAAATTAATGTTGAGGATGTTGCCAATGCAGCTCTTTATCTGGCTAGT 723  
DB 676 CTGAAAGTCCAAAATTAATGTTGAGGATGTTGCCAATGCAGCTCTTTATCTGGCTAGT 735

QY 724 GATGAGGCAAAATACGTGAGTGACACAACTCTGTTTCATTGATGGAGGTTTCAGGCTCTGC 783  
DB 736 GATGAGGCAAAATACGTGAGTGACACAACTCTGTTTCATTGATGGAGGTTTCAGGCTCTGC 795

QY 784 AATTCTGTAATCAAGTGTTCCTCAATATCCAGATTCT 819  
DB 796 AATTCTGTAATCAAGTGTTCCTCAATATCCAGATTCT 831

## RESULT 4

AAZ45734  
ID AAZ45734 standard; cDNA; 828 BP.

XX AC  
XX AC  
XX AC

XX AC  
XX AC  
XX AC

XX AC

XX AC  
XX AC  
XX AC

XX AC  
XX AC  
XX AC

XX AC  
XX AC  
XX AC

XX AC  
XX AC  
XX AC

XX AC  
XX AC  
XX AC

XX AC  
XX AC  
XX AC

XX AC  
XX AC  
XX AC

KW enterolactone; enterodiol; neutriceutical; dietary supplement;  
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.  
XX Forsythia intermedia.  
OS  
XX Key Location/Qualifiers  
FH 1-828  
FT CDS  
FT  
FT  
FT  
FT  
FT  
FT  
PN WO9955846-A1.  
XX  
XX 04-NOV-1999.  
PD  
XX 23-APR-1999; 99WO-US08975.  
XX 24-APR-1998; 98US-0082977.  
XX (UNIW ) UNIV WASHINGTON STATE RES FOUND.  
XX Xia Z, Costa MA, Davin LB, Lewis NG;  
PI  
DR WPI: 2000-126356/11.  
DR P-PSDB; AAY54416.  
XX  
XX New nucleic acid molecule encoding an enzyme involved in lignan  
XX biosynthetic pathway, useful for producing large amounts of lignans  
XX  
XX Claim 9; Page 55-56; 66pp; English.  
XX  
CC The present sequence encodes a secoisolariciresinol dehydrogenase  
CC protein. The enzyme is involved in the lignan biosynthetic pathway.  
CC The secoisolariciresinol dehydrogenase proteins have a molecular  
CC weight of 27.31 kDa, an isoelectric point of 5.9-6.85, and require  
CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase  
CC nucleic acids are used for the recombinant expression of the enzymes.  
CC It is also used to obtain expression or enhanced expression of  
CC secoisolariciresinol dehydrogenase in plants or to alter lignan  
CC biosynthesis. The enzyme is used for production of the pharmacologically  
CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase  
CC proteins and nucleic acids can be utilized to: elevate or otherwise  
CC alter the levels of health-protecting lignans, including phytoestrogens  
CC such as enterolactone and enterodiol, in plant species, including  
CC vegetables, grains and fruits and to food items incorporating material  
CC derived from such genetically altered plants; genetically alter plant  
CC species to provide an abundant, natural supply of lignans useful for  
CC a variety of purposes, for example as nutraceuticals and dietary  
CC supplements; to genetically alter living organisms to produce an  
CC abundant supply of optically pure lignans having desirable biological  
CC properties, for example (-)-trachelogenin which possesses antiviral  
CC properties, and (-)-podophyllotoxin.  
XX  
XX Sequence 828 BP; 235 A; 172 C; 196 G; 225 T; 0 other;  
SQ

Query Match 89.9%; Score 736; DB 21; Length 828;  
Best Local Similarity 93.9%; Pred. No. 8.8e-206;  
Matches 766; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 4 CAGCTTCCGAATGCAATGCGAAGAGGCTAGAGGAAAAGTTGCCCTTATAACAGGAGGA 63  
DB 13 CAGCTTCCGAATGCAATGCGAAGAGGCTAGAGGAAAAGTTGCCCTTATAACAGGAGGA 72

QY 64 GCACGTGAATGGAGAAACACACAGCAAACTCTCTCCCAACATGGAGCCAAAGTTGCC 123  
DB 73 GCACGTGAATGGAGAAACACACAGCAAACTCTCTCCCAACATGGAGCCAAAGTTGCC 132

QY 124 ATTGCTGATGCCAAGATGAATAGTCACTAGTGTGCGAGGCCATTGGGCACATTCCTCAAT 183  
DB 133 ATTGCTGATGCCAAGATGAATAGTCACTAGTGTGCGAGGCCATTGGGCACATTCCTCAAT 192

QY 184 TCCACCTACATCCACTGTGATGTTACTAATGAAGACGGTGTAAATAATGCCGTGGACAAC 243  
DB 184 TCCACCTACATCCACTGTGATGTTACTAATGAAGACGGTGTAAATAATGCCGTGGACAAC 243





RESULT 7  
AAZ45733  
ID AAZ45733 standard; cDNA; 831 BP.  
XX  
AC AAZ45733;  
XX  
DT 06-APR-2000 (first entry)  
XX  
DE Secoisolaricresinol dehydrogenase cDNA clone SMDEHY511.  
XX  
KW Secoisolaricresinol dehydrogenase; lignan biosynthetic pathway;  
KW lignan; matairesinol; health-protecting lignan; phytoestrogen;  
KW enterolactone; enterodiol; neutriceutical; dietary supplement;  
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.  
XX  
OS Forsythia intermedia.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 1..831  
FT /tag= a  
FT /product= "secoisolaricresinol dehydrogenase"  
FT /note= "no termination codon given"  
XX  
XX WO955846-A1.  
XX  
XX 04-NOV-1999.  
XX  
XX 23-APR-1999; 99WO-US08975.  
XX  
XX 24-APR-1998; 98US-0082977.  
XX  
XX (UNIW ) UNIV WASHINGTON STATE RES FOUND.  
XX  
XX Xia Z, Costa MA, Davin LB, Lewis NG;  
XX  
XX WPI; 2000-126356/11.  
XX P-PSDB; AAY54415.  
XX  
XX New nucleic acid molecule encoding an enzyme involved in lignan  
XX biosynthetic pathway, useful for producing large amounts of lignans -  
XX  
XX Claim 9; Page 52-54; 66pp; English.  
XX  
XX The present sequence encodes a secoisolaricresinol dehydrogenase  
XX protein. The enzyme is involved in the lignan biosynthetic pathway.  
XX The secoisolaricresinol dehydrogenase proteins have a molecular  
XX weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require  
XX NAD or NADP as a cofactor. The secoisolaricresinol dehydrogenase  
XX nucleic acids are used for the recombinant expression of the enzyme.  
XX It is also used to obtain expression or enhanced expression of  
XX secoisolaricresinol dehydrogenase in plants or to alter lignan  
XX biosynthesis. The enzyme is used for production of the pharmacologically  
XX active lignan, matairesinol. The secoisolaricresinol dehydrogenase  
XX proteins and nucleic acids can be utilized to: elevate or otherwise  
XX alter the levels of health-protecting lignans, including phytoestrogens  
XX such as enterolactone and enterodiol, in plant species, including  
XX vegetables, grains and fruits and to food items incorporating material  
XX derived from such genetically altered plants; genetically alter plant  
XX species to provide an abundant, natural supply of lignans useful for  
XX a variety of purposes, for example as neutriceuticals and dietary  
XX supplements; to genetically alter living organisms to produce an  
XX abundant supply of optically pure lignans having desirable biological  
XX properties, for example (-)-trachelogenin which possesses antiviral  
XX properties, and (-)-podophyllotoxin.  
XX  
SQ Sequence 831 BP; 244 A; 181 C; 195 G; 211 T; 0 other;  
Query Match 86.9%; Score 712; DB 21; Length 831;  
Best Local Similarity 92.0%; Pred. No. 9.8e-199;  
Matches 751; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
QY 4 CAGCTTCGAACATGCTTCGAAAGAGGCTAGAGGAAAGTTGCCCTTATAACAGGAGGA 63

Db 16 CAGGTTCTAACTGCAATCAAGAAGGCTAGAGGAAAGTTGCCCTTATAACAGGAGGA 75  
QY 64 GCAGTGGAAATGGAGAAACACAGCAAAACTCTTCTCCCAACATGAGCCAAAGTTGCC 123  
Db 76 GCAGTGGAAATGGAGAAATTCACAGCAAAACTCTTCTCCCAACATGAGCCAAAGTTGCC 135  
QY 124 ATTGCTGATGTCCAAGATGAATTAGGTCACTCAGTTGTGCGAGGCCATTGGCACTTCCAAT 183  
Db 136 ATTGCTGATGTCCAAGATGAATTAGGTCACTCAGTTGTGCGAGGCCATTGGCACTTCCAAT 195  
QY 184 TCACCTACATCCACTGTGATGTTTACTAATGAAGAGCGGTGTTAAAAATGCCGTGGACAAC 243  
Db 196 TGCATATCATCCACTGCGATGTTTACCAATGAAGACGATGTTAAAAATGCCGTGGACAAC 255  
QY 244 ACAGTTTCAACCTATGGAAGAACTGGACATTATCTTCAGCAATGSCAGAAATTTCTGTATCCC 303  
Db 256 ACAGTTTCAACCTATGGAAGAACTGGACATTATCTTCAGCAATGSCAGAAATTTCTGTATCCC 315  
QY 304 AACAGCCCCGCATCATAGACAACGAAAAAGCAGACTTTGNAAGCGCTTCTCAGTGTAAAT 363  
Db 316 AACAGCCCCGCATCGTAGACAACGAAAAAGCAGACTTTGNAAGCGCTTCTCAGCGTAAAT 375  
QY 364 GTAAACCGGAGTTTCTCTATGTCATGAAGCAGCAGCAGTGTTATGATTTCCAGCAGCCAGT 423  
Db 376 GTAAACCGGTTTCTCTATGTCATGAAGCAGCAGCAGTGTTATGTTGGTGGCAGCAGCAGT 435  
QY 424 GGCAACATAAATTTCCACTGCTAGTTTAAAGCTCAACTATATGGTGGTGGTTCACATGCC 483  
Db 436 GGCAGCATAAATTTCCACTGCTAGCTAAAGCTCAACAAATTTGGTGGTGGTTCACATGCT 495  
QY 484 TATTTGGTTCAAAGCATGCTGTGTAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGA 543  
Db 496 TATTTGGTTCAAAGCATGCTGTGTAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGA 555  
QY 544 CAATTTGGCATTAGGTTAATTTGTTGTTCTCTTCCGGGCTTCCCTACGGCTTTTAGGCAAG 603  
Db 556 CAATTTGGCATTAGGTTAATTTGTTGTTCTCTTACGGCTTCTACGGCTTTTAGGCAAG 615  
QY 604 AAATTTTCAGGAGTAAAAATGAAGAAGAAATTTGACAATGTAATAAATTTTCGGGAAAT 663  
Db 616 AAATTTGTAGGCTTGAATAATGACGAAGATTTGGAGAAATGCAATGAGCCCTTATGGGAAAT 675  
QY 664 TTGAAAGTCCAAAATTTAATGTTGAGGATTTGCCAATGCAGCTCTTTATCTGGCTAGT 723  
Db 676 CTGAAAGGTACAAATTTGAGGCTGAGGAGCTGCGCAATGCAGCTCTTTATCTGGCAAGT 735  
QY 724 GATGAGGCAAAATACGTGAGTGAGACACAATCTGTTCAATGATGGAGGGTTACGCGTCTGC 783  
Db 736 GATGAGGCAAAATATGTGAGTGAGACACAATCTGTTCAATGATGGAGGGTTACGCGTCTAC 795  
QY 784 AATTCGTGAATCAAAAGTGTTCCAATATCCAGATTCT 819  
Db 796 AATTCGTGAATCAAAATGTTCCAATATCCAGACT 831  
RESULT 8  
AAC40811  
ID AAC40811 standard; DNA; 1120 BP.  
XX  
AC AAC40811;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 29624.  
XX  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
XX Arabidopsis thaliana.  
XX  
PN EP1033405-A2.

XX	06-SEP-2000.	PR	02-JUL-1999;	99US-0142055.
PD		PR	06-JUL-1999;	99US-0142390.
XX		PR	08-JUL-1999;	99US-0142803.
PF		PR	09-JUL-1999;	99US-0142920.
XX	25-FEB-2000; 2000EP-0301439.	PR	12-JUL-1999;	99US-0142977.
XX		PR	13-JUL-1999;	99US-0143342.
PR	25-FEB-1999;	PR	14-JUL-1999;	99US-0143624.
PR	05-MAR-1999;	PR	15-JUL-1999;	99US-0144005.
PR	09-MAR-1999;	PR	16-JUL-1999;	99US-0144085.
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144086.
PR	25-MAR-1999;	PR	19-JUL-1999;	99US-0144325.
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144331.
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144332.
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144333.
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144334.
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144335.
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144352.
PR	19-APR-1999;	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086.
PR	04-MAY-1999;	PR	21-JUL-1999;	99US-0145088.
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145085.
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145087.
PR	07-MAY-1999;	PR	22-JUL-1999;	99US-0145089.
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145192.
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145145.
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145913.
PR	18-MAY-1999;	PR	27-JUL-1999;	99US-0145918.
PR	19-MAY-1999;	PR	27-JUL-1999;	99US-0145919.
PR	20-MAY-1999;	PR	28-JUL-1999;	99US-0145951.
PR	21-MAY-1999;	PR	28-JUL-1999;	99US-0146386.
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	PR	02-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	PR	03-AUG-1999;	99US-0147038.
PR	28-MAY-1999;	PR	04-AUG-1999;	99US-0147204.
PR	01-JUN-1999;	PR	04-AUG-1999;	99US-0147302.
PR	03-JUN-1999;	PR	05-AUG-1999;	99US-0147192.
PR	07-JUN-1999;	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	PR	11-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	PR	12-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	PR	12-AUG-1999;	99US-0148341.
PR	17-JUN-1999;	PR	12-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	PR	20-SEP-1999;	99US-0154779.





QY	7	CTTCGA	ACTGC	ATTCG	CAAGA	AGGCT	AGA	GAAAA	AGTTT	GCCTT	TAT	AAC	AGG	AGG	AGCC	66
Db	47	CTTC	CAAT	CAC	CAAT	AGC	CAAA	AGG	TAA	AGG	TAA	AGG	TAA	AGG	TAA	106
QY	67	AGT	GGA	ATT	TGG	AAAC	CCAC	GAA	AACT	CTT	CTC	CC	AC	ATG	GG	126
Db	107	AGC	GC	AT	AG	CGCC	CCAC	AG	CTAG	CTTT	TGT	T	CA	CA	ATG	166
QY	127	GCT	GAT	G	TCC	AA	GAT	GA	T	TAG	T	AG	T	CA	CT	186
Db	167	CG	AG	CA	TT	CA	AG	CA	AC	CT	CT	T	T	AG	T	226
QY	187	AC	CT	AC	AT	CC	ACT	GT	GA	T	T	AC	T	AA	AA	246
Db	227	AT	CT	TT	AT	CC	AT	TC	AA	T	T	G	T	G	T	286
QY	247	G	TT	CA	AC	CT	AT	G	GA	AA	CT	GG	CA	AT	T	306
Db	287	AT	T	G	CC	AA	TT	T	G	TA	AG	CT	GG	CA	AT	346
QY	307	AG	G	C	C	C	C	C	AT	C	AT	C	AT	C	AT	366
Db	347	AT	T	T	C	C	AG	AT	C	T	T	AG	AT	C	T	403
QY	367	AC	G	G	AG	T	T	T	C	CT	AT	GC	AT	GC	AT	425
Db	404	G	T	T	G	C	G	C	T	T	C	T	AT	GC	AT	463
QY	426	CA	AC	AT	AT	T	CC	AC	T	CT	AG	T	T	AA	AG	485
Db	464	T	T	C	CA	AT	CT	CA	CA	CA	AG	T	AT	T	G	503
QY	486	T	T	G	G	T	T	CA	AG	CA	-	TG	C	T	AG	544
Db	504	T	T	C	AT	CA	AA	AG	G	T	AG	T	T	CC	AA	563
QY	545	A	A	T	T	G	C	AT	T	A	G	G	T	T	A	604
Db	564	A	G	T	A	G	G	A	A	A	T	T	A	A	A	623
QY	605	A	A	T	T	T	C	AG	G	A	T	T	A	A	A	664
Db	624	A	T	G	C	C	T	T	G	A	A	T	T	A	A	683
QY	665	T	G	A	A	G	T	C	C	A	A	A	A	A	A	724
Db	684	T	G	A	A	G	G	A	G	C	T	T	A	G	A	742
QY	725	A	T	G	G	C	A	A	A	T	T	A	A	A	A	784
Db	743	A	T	G	A	T	T	C	T	A	A	T	T	C	T	802
QY	785	A	T	786												
Db	803	A	T	804												
RESULT 10																
AAC38191																
ID	AAC38191 standard; DNA; 1136 bp.															
XX																
AC	AAC38191;															
XX																
DT	17-OCT-2000 (first entry)															
XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 20099.															
DE																
KW	Hybridisation assay; genetic mapping; gene expression control;															
KW	protein identification; signal transduction pathway;															
XX	metabolic pathway; promoter; termination sequence; ss.															
OS	Arabidopsis thaliana.															
XX																

PN	EP1033405-A2.
XX	
PD	
XX	06-SEP-2000.
XX	
XX	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123348.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134236.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.
PR	20-MAY-1999; 99US-0135124.
PR	21-MAY-1999; 99US-0135353.
PR	24-MAY-1999; 99US-0135629.
PR	25-MAY-1999; 99US-0136021.
PR	27-MAY-1999; 99US-0136392.
PR	28-MAY-1999; 99US-0136782.
PR	01-JUN-1999; 99US-0137222.
PR	03-JUN-1999; 99US-0137528.
PR	04-JUN-1999; 99US-0137502.
PR	07-JUN-1999; 99US-0137724.
PR	08-JUN-1999; 99US-0138094.
PR	10-JUN-1999; 99US-0138540.
PR	10-JUN-1999; 99US-0138847.
PR	14-JUN-1999; 99US-0139119.
PR	16-JUN-1999; 99US-0139452.
PR	17-JUN-1999; 99US-0139453.
PR	18-JUN-1999; 99US-0139454.
PR	18-JUN-1999; 99US-0139455.
PR	18-JUN-1999; 99US-0139456.
PR	18-JUN-1999; 99US-0139457.
PR	18-JUN-1999; 99US-0139458.
PR	18-JUN-1999; 99US-0139459.
PR	18-JUN-1999; 99US-0139460.
PR	18-JUN-1999; 99US-0139461.
PR	18-JUN-1999; 99US-0139462.
PR	18-JUN-1999; 99US-0139463.
PR	18-JUN-1999; 99US-0139750.
PR	21-JUN-1999; 99US-0139763.
PR	21-JUN-1999; 99US-0139817.
PR	22-JUN-1999; 99US-0139809.
PR	23-JUN-1999; 99US-0140353.
PR	23-JUN-1999; 99US-0140354.
PR	24-JUN-1999; 99US-0140695.
PR	28-JUN-1999; 99US-0140823.
PR	29-JUN-1999; 99US-0140991.
PR	30-JUN-1999; 99US-0141287.
PR	01-JUL-1999; 99US-0141842.

```
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0145386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.

PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0158458.
PR 29-SEP-1999; 99US-0158596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 22.8%; Score 187; DB 21; Length 1136;
Best Local Similarity 54.9%; Pred. No. 1.7e-44;
Matches 420; Conservative 0; Mismatches 330; Indels 15; Gaps 2;

QY 23 CAAGAAGGCTAGAGCAAAAGTTGCCCTTATACAGAGAGAGCCAGTGGAAATGGAGAA 82
Db 294 CAAGAAGGCTAGAGCAAAAGTTAGGTAGCAGTATCATCTGGAGAGCAAGTGGGAATGGCAAG 353
QY 83 CCACAGCAAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTTGCTGTGATGTCCAAGATG 142
Db 354 CAACAGCCGGAATTCATCAGTCATGGAGCCAAAGTTATCATTTGCCGATATCCACCCG 413
QY 143 AATTAGTCACTCAGTTGTGCGAGGCCATTTGGCAGTTCCTCAATTCACCTACATCCACTGTG 202
Db 414 AGATTGGGC---GAGAAACCGGAGAACTCGGTCCAGTTGTGCTTACTTCCCATGCG 470
QY 203 ATGTTTACTAATGAAGACGGTGTAAATAATGCCGTGGACACACAGTTTCAACCTATGGAA 262
Db 471 ATGTGACCAAGAATCAGACATTTGCTAACGCAAGTTGACTTCGCTGTCTCGTCCATACAA 530
QY 263 AACTGGACATATGTTTCAGCAATGCAGGAATTTCTGATCCCAACAGCCCGCCATCATAG 322
Db 531 AGCTCGACATATGTACACAATGCTGGTATTCCTTGCAAAAGCCCTCTCTAGTATCGTTG 590
QY 323 ACAACGAAAAAGCAGACTTTTGAACGGCTTCTCAGTGTAAATGTAACCGGAGTTTCTCTAT 382
Db 591 ATCTTGATCTCAATGTTTTCGACAAGTAAACAACAAATGTCCTGGAGTTCATGGCAG 650
QY 383 GCATGAAGCACGACGACGCTGTTATGATTCACGACGAGTGGCAACATAATTTCCACTG 442
Db 651 GAATCAACATGCTGCTGCTGATGATCCCGCGTAACCTCTGGATCCATCATTTGTGCGAG 710
```

```
QY 443 CPAGTTTAAAGCTCAACATATGGGTGGTTCCTTCACATCCCTATTGTTGTTCAAGCAATG 502
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 711 GGAGTGTACCGGATGATGGCGGTTTACGCCAACATACCTTACACGCGTCTCAAAATCCG 770
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 503 CTGTTTAGCCCTTACTAGGAATCTGGCAGTCCGAGCTCGGACAATTTGGCAATTAGGGTTA 562
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 771 CTGTTATCGGAATTGTAAAGATCAACAGCTTCAGCAACTATGCAAGCACAGGATCCGGTCA 830
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 563 ATTGTTTCTCCTTTTCGGGCTTCCCTACGGCTTTTAGGCAAGAAAT-----TTT 610
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 831 ACTGCAATTCCTCTTTTGGCATCAACAATCATCTGATGATGAGATCGCACAGATTT 890
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 611 CAGGCAATTAATAAGACAAGAAATTTGAGAAATGTAATAAATTTGCGGAAATTTGAAG 670
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 891 ACCCGGTGTTGATGATCTCAGGCTGATCCAGATATGCGAGTACAGAGTGTTAAATG 950
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 671 GTCCAAAATTTAATGTTGAGAGTGTGGCCAAATCGACGCTTTTATCTGCGTAGTGATGAGG 730
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 951 GAGAGTTTGTGAACCAACCGATGTAGCTAATGACGCGGTGATCTCGCTTCCGATGATT 1010
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 731 CAAAATAGCTGAGTGGACACAATCTGTTTCATTGATGAGGGGTTCA 775
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 1011 CAAAGTATGTAATGGGCATAATCTGTTGATGAGGAGATTCA 1055
    |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 11
AAC43068
ID AAC43068 standard; DNA; 852 BP.
XX
AC AAC43068;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37909.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
```







PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	07-MAY-1999;	99US-0132487.
PR	11-MAY-1999;	99US-0132863.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	21-MAY-1999;	99US-0135124.
PR	24-MAY-1999;	99US-0135299.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137702.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139450.
PR	21-JUN-1999;	99US-0139763.
PR	22-JUN-1999;	99US-0139817.
PR	23-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	24-JUN-1999;	99US-0140354.
PR	28-JUN-1999;	99US-0140695.
PR	29-JUN-1999;	99US-0140823.
PR	30-JUN-1999;	99US-0140991.
PR	01-JUL-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	02-JUL-1999;	99US-0142154.
PR	06-JUL-1999;	99US-0142055.
PR	09-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	12-JUL-1999;	99US-0142920.
PR	13-JUL-1999;	99US-0142977.
PR	14-JUL-1999;	99US-0143424.
PR	15-JUL-1999;	99US-0143624.
PR	16-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	19-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	21-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	03-AUG-1999;	99US-0146389.
PR	04-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	05-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147535.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149923.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149002.
PR	23-AUG-1999;	99US-0149920.
PR	23-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	14-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.



PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match 18.3%; Score 150.2; DB 21; Length 774;			
Best Local Similarity 52.7%; Pred. No. 8.9e-34;			
Matches 401; Conservative 0; Mismatches 348; Indels 12; Gaps 3;			
QY	28	AGCGTAGAGGAAGTCCCTTATACAGGAGGAGCCAGTGGATTCGAGAAACCA 87	
Db	13	AGATTGATGGCAAAATCGCAATATAACAGCGGAGTAGCGGGATTGGAGCGGAAGCG 72	
QY	88	GCAAAACTCTTCCCAACATGGAGCCAAAGTTGCCATTGCTGATGTCACAGATGAATTA 147	
Db	73	GTTAGGCTGTTACGCGACCCAGGAGTAGGTGTCATCGTTGACTTTCAGAGAGACTT 132	
QY	148	GGTCACTAGTTGTCGAGGCCATTTGCGACTTCCAAATTCACCTACATCCACTGTGATGT 207	
Db	133	GGTCAAAACGTTGCCGTTCTGTCGGGAAGACAAAGCAAGTTTTTACCGTTCCGATGT 192	
QY	208	ACTAANTGAACAGCGTCTTAAATTCGTCGTCGACCAACACAGTTTCAACCTATGCAAACTG 267	
Db	193	ACAAAGCAAGGAGGTAGAAAACGCGTTAGTTTACCGTCGAAAAGTACGGGAAGCTT 252	
QY	268	GACATTATGTTACGAATCAGGAATTTCTGATCCCAACAGGCCCGCATCATAGACAAC 327	
Db	253	GACGTTCTCTTAGTAACCGCGCTTATGGAAC---AGCCGGGAAGCTTCTCGACTTG 309	
QY	328	GAAGAACGAGACTTTGAACGCTTCTCAGTGTAAATGTAAACGGAGTTTTCCTATGCATG 387	
Db	310	AATCTGGAACAGTTTGACCGCAACCATGGCGGTCAACGTTTCGTGGTGGCTGCGTTTATC 369	
QY	388	AAGCAGCAGCAGCTGTTATGATTCAGCACCA---GTGGCAACATAATTTCCACTGCT 444	
Db	370	AAACGCTGCACGACCCATGGTGGAGAAAGGACCGCGGTGTCATCGTATGTACGACC 429	
QY	445	AGTTTAAGCTCAACTATGGGTGGTGTCTTTCACATGCCTATTGTTGTTCAAGCATGCT 504	
Db	430	AGCGTCGCTCGAGATCGTGTGTCAGGACCTACGCGTACACGCGTCTAAGCAGCT 489	
QY	505	GTGTTAGCCCTTACTAGGAATTCGGCAGTCGAGCTCGGACAAATTTGGCATTAGGGTTAAT 564	
Db	490	CTTCTCGGCTGTTTAAATCGGCTTGTGGCGGCTAGGGAAGTACGGGATTAGAGTCAAC 549	
QY	565	TGTTTGTCTCTTCGGCTTCTACGCGTTTAGGCAAGAAATTTTCAGGGATTAAAT 624	
Db	550	GGCGTTGCACCATACCGGTGGCGGCGGCGATA-----AATAGCGTGACGAGGAACG 603	
QY	625	GAAGAGAATTTTGAGATGTAATAAACTTTGCGGGAATTTGAAAGGTCCAAAATTTAAT 684	
Db	604	GTGAGGATGGTGAGGAATATACGCGACGCCAGGGGATTTCTAAAGGTGTGGTCTTAG 663	
QY	685	GTTGAGATGTTGCCAATGACGCTTTTATCTGGCTAGTGATGAGCAAAATACGTGACT 744	
Db	664	GCTCGCCATGTGCGAGAGCGGCTTGTCTTCTGCTTCGGATGATTCGGCTTACGTTAGC 723	
QY	745	GGACAAATCTGTTATTCATGAGGGGTTACGGCTTCGAA 785	
Db	724	GTCAGAAATCTGGCTTTCAGCGGGGTTATAGCGTGTAA 764	

RESULT 14

ABN98918/c

ID ABN98918 standard; DNA; 833 BP.

XX AC ABN98918;

XX DT 01-AUG-2002 (first entry)

XX Arabidopsis thaliana expressed polynucleotide SEQ ID NO 686.

XX Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;

KW disease; crop; thale cress; tolerance factor; insect; pathogen;

KW nutrition; ds.

XX Arabidopsis thaliana.

XX US2002023281-Al.

XX PD 21-FEB-2002.

XX PF 26-JAN-2001; 2001US-0770445.

XX PR 27-JAN-2000; 2000US-178472P.

XX (GORL/) GORLACH J.

PA (ANY/) AN Y.

PA (HAMI/) HAMILTON C M.

PA (PRIC/) PRICE J L.

PA (RAIN/) RAINES T M.

PA (YUY/) YU Y.

PA (NAME/) RAMEAKA J G.

PA (PAGE/) PAGE A.

PA (MATH/) MATHAW A V.

PA (LEDF/) LEDFORD B L.

PA (WOES/) WOESSNER J P.

PA (HAAS/) HAAS W D.

PA (GARC/) GARCIA C A.

PA (KRIC/) KRICKER M.

PA (SLAT/) SLATER T.

PA (DAVI/) DAVIS K R.

PA (ALLE/) ALLEN K.

PA (HOFF/) HOFFMAN N.

PA (HURB/) HURBAN P.

XX Goriach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;

PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;

PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;

PI Hurban P;

XX WPI; 2002-400781/43.

XX New Arabidopsis thaliana nucleic acid for identifying homologous genes,

PT producing compositions that modulate the expression or function of its

PT encoded protein, and mapping functional regions of protein -

XX Claim 1; SEQ ID NO 686; 49pp + Sequence Listing; English.

XX The invention relates to an Arabidopsis thaliana nucleic acid (I)

CC comprising a sequence capable of hybridising under stringent conditions

CC to a sequence selected from any one of 999 sequences (ABN98233-ABN99231),

CC given in the specification or its fragment. A polypeptide (II) encoded by

CC (I), a transgenic plant (III) comprising an exogenous nucleic acid or a

CC genetically modified cell (IV) comprising an exogenous nucleic acid, is

CC useful for screening a candidate agent for its biological effect. (I) is

CC useful in identifying homologous or related genes, in producing

CC compositions that modulate the expression or function of its encoded

CC protein, mapping functional regions of the protein and in studying

CC associated physiological pathways. (I) is also useful for the genetic

CC manipulation of cells, particularly plant cells. (I) is also useful in

CC screening assays of various plant strains to determine the strains that

CC are best capable of withstanding a particular disease or environmental

CC stress. (II) and (III) are useful for screening of biologically active  
 CC agents, e.g. fungicides, insecticides, etc., for elucidating biochemical  
 CC pathways. The screened agents are useful in improved methods of treating  
 CC crops to prevent or treat disease. (II) are also useful in screening  
 CC programs to identify agents that mimic or enhance the action of tolerance  
 CC factors. Such agents are useful in improved methods of treating crops to  
 CC enhance their tolerance to environmental stress. (I) is also useful  
 CC for enhancing or inhibiting production of a biosynthetic product in a  
 CC plant. (III) is useful for identifying other mediators that may induce  
 CC expression of proteins of interest, for establishing the extent to which  
 CC any specific insect and/or pathogen is responsible for damage to a  
 CC particular plant, for identifying other mediators that enhance or induce  
 CC tolerance to environmental stress, for identifying factors involved in  
 CC biosynthetic pathways of nutritional, commercial, or medicinal value and  
 CC for identifying productions of nutritional, commercial or medicinal  
 CC value. (IV) is useful in the study of genetic function and regulation,  
 CC that alteration of the cellular metabolism and for screening compounds  
 CC that may affect the biological function of the gene or gene products.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from USPTO  
 CC at seqdata.uspto.gov/sequence.html?DocID=99999770445.

XX at seqdata.uspto.gov/sequence.html?DocID=99999770445.  
 SQ Sequence 833 BP; 193 A; 250 C; 177 G; 203 T; 10 other;  
 Query Match 18.2%; Score 149; DB 24; Length 833;  
 Best Local Similarity 52.2%; Pred. No. 2.1e-33;  
 Matches 393; Conservative 0; Mismatches 350; Indels 12; Gaps 3;

QY 32 TAGAAGAAAGTTCCTTATACAGGAGGAGCCGATGGAAATGGAGAAACACAGCAA 91  
 DB 833 TGGATGCAAAATCGCAATATACAGGCGAGCTAGCGGATGGAGCGGAGCGGTTA 774  
 QY 92 AACTCTTCTCCCAATGAGCAAGTTCGATGTCGATGTCCTCAAGATGAATAGGTC 151  
 DB 773 GGCTGTCAGGACACGAGCGTAGTGGTATCGTGTGACTTTCAGAGAAAGACTTGGTC 714  
 QY 152 ACTCAGTGTGCGAGCCATGCGCATTCATTCACCTACATCCACTGATGTTACTA 211  
 DB 713 AAAACGTTGCGGTTCTGTCGGAAGAACACAGCANNNTTACCGTGGGATGTACAA 654  
 QY 212 ATGAAGAGCGGTGTAATAATCCGTTGACACACAGTTCACCTATGGAACACTGGACA 271  
 DB 653 ACGAAAGAGGTNNAAACCGTGAATGACCTGCGTGAAGAGCGGATGAGCG 594  
 QY 272 TTATGTCAGCAATGAGGAATTCATGATCCCAACAGCCCGCATACAGCAAGAA 331  
 DB 593 TTCTCTTTNNAAACCGCGGTATGGAAC---AGCGGGAAGCTTCGCGATGAATC 537  
 QY 332 AAGCAGACTTTGAACGCGTCTCAGTGTAAATGTAACCGAGTTTCCATGATGATGAGC 391  
 DB 536 TGGACAGTTTCAACGACCATGGTGGAGAAAGCAGCGTGGTCAATCGTATGACGACGG 477  
 QY 392 ACGCAGCGTGTATGATTCACGACGCA---GTGGCAACATAATTCACGCTAGTT 448  
 DB 476 ACGTGCACGAGCCATGGTGGAGAAAGCAGCGTGGTCAATCGTATGACGACGG 417  
 QY 449 TAAGTCAACTATGGTGGTGGTCTTTCATGCTGCTATGCTTCAAGCATGCTGTGT 508  
 DB 416 TGGCTCGGAGATCGGTGGTCCAGGACCTCAAGGATACAGCGGCTAAGCAGCTCTTC 357  
 QY 509 TAGCCCTTACTAGGAATTCGACGCTGAGCTCGGCAATTTGCGATAGGTTAATGTT 568  
 DB 356 TCGGCTGGTAAATTCGCTTGTGGCGGGTACGGAGTACGGGATAGTCAACGCG 297  
 QY 569 TGTCTCTTTCCGGGTTCTTACGGCTTTAGCGAAGAAATTTTCAGGATTAAGATGAAG 628  
 DB 296 TTGCACCATAGCGGTGGCGAGCGGATA-----AATAGCGGTGACGAGAAACGGTGA 243  
 QY 629 AAGAAATTTGAGAAATTAATAAATTTTCGGGAAATTTGAAGAGTCCAAATTAATGTTG 688  
 DB 242 GGATGGTGGAGGAATATAGCGACCGCGGATTTCTCAAGAGTGTGCTTAAGGCTC 183  
 QY 689 AGGATGTTGCCAATGCAGCTTTTATCTGCTAGTGTAGTCAAGCAAAATACGTTAGTGCAC 748

DB 182 GCATGTGGCAGAGCGGCTTTGTTCTGCTCGGATGATTCGGTTACGTTAGCGGTC 123  
 QY 749 ACAATCTGTTCAATGATGAGGGTTTCAGCGCTCTGCAA 785  
 DB 122 AGAATCTGCTGTTACACGGGGTTATAGCGTCGTAA 86  
 RESULT 15  
 AAC50609  
 ID AAC50609 standard; DNA: 774 BP.  
 XX AAC50609;  
 AC AAC50609;  
 DT 18-OCR-2000 (first entry)  
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 65475.  
 XX Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway;  
 KW metabolic pathway; promoter; termination sequence; ss.  
 XX Arabidopsis thaliana.  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 XX 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 04-MAY-1999; 99US-0132407.  
 PR 05-MAY-1999; 99US-0132484.  
 PR 06-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 07-MAY-1999; 99US-0132487.  
 PR 11-MAY-1999; 99US-0132863.  
 PR 14-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 18-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 20-MAY-1999; 99US-0134941.  
 PR 21-MAY-1999; 99US-0135124.  
 PR 24-MAY-1999; 99US-0135353.  
 PR 25-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 15:19:44 ; Search time 1467 Seconds  
(without alignments)  
9041.652 Million cell updates/sec

Title: US-09-673-918A-1

Perfect score: 819

Sequence: 1 atgcagcttcgaactgcatt.....tggccaataatccagattct 819

Scoring table: IDENTITY\_NUC

Gapop 10\_0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_nam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	362.4	44.2	738	14	BQ870873
2	348.4	42.5	704	14	BQ875148
3	342.6	41.8	702	14	BQ873587
4	290.6	35.5	585	12	BE920368
5	288.4	35.2	676	12	BG350846
6	271.6	33.2	717	14	BM812901

7	258.4	31.6	637	14	BQ630495
8	258.4	31.6	686	10	BE205431
9	257.6	31.5	682	12	BQ456805
10	257.4	31.4	567	13	BI701214
11	257.4	31.4	681	12	BE521399
12	256.6	31.3	695	14	BQ862029
13	254.6	31.1	655	10	AW686040
14	252.6	30.8	544	14	BQ982959
15	252.6	30.8	544	14	BQ982959
16	248.4	30.3	704	10	BQ914508
17	245.8	30.0	621	12	BQ344207
18	240.4	29.4	617	14	BG134389
19	239.8	29.3	559	10	BQ988267
20	239.4	29.2	682	10	AW202061
21	238.6	29.1	637	12	AW185846
22	238	29.1	574	14	BQ080788
23	237.6	29.0	745	13	BM110645
24	237.2	29.0	590	10	AW278238
25	236.2	28.8	897	10	BE660446
26	235	28.7	804	14	BQ120237
27	229.8	28.1	620	10	BE556284
28	229	28.0	552	13	BI892743
29	228.4	27.9	743	14	BQ046202
30	228.2	27.9	574	10	AW933981
31	227.6	27.8	585	10	BE187511
32	222.6	27.2	767	12	BG523029
33	220.2	26.9	662	12	BF641439
34	219.4	26.8	581	13	BM093608
35	216	26.4	721	10	BE660445
C	36	214.6	744	12	BE822419
	37	214.4	598	10	AW689744
	38	212.4	586	10	AW57208
	39	210.6	715	9	AJ497310
C	40	209	767	10	AW349377
	41	207	578	12	BF004132
	42	206	514	12	BE999230
	43	206	536	12	BG448958
	44	206	641	13	BM111970
	45	205.2	624	14	BQ875438

ALIGNMENTS

RESULT 1	BQ870873	BQ870873	BQ870873	738 bp	mRNA	linear	EST 15-AUG-2002
LOCUS	BQ870873	QG110E14.y9.abl	QG-ABCDI	lettuce	salinas	Lactuca sativa	cDNA clone
DEFINITION	BQ870873	QG110E14	mRNA sequence.				
ACCESSION	BQ870873	BQ870873.1	GI:22257413				
VERSION	BQ870873	EST					
KEYWORDS	BQ870873	Lactuca sativa.					
SOURCE	BQ870873	Lactuca sativa					
ORGANISM	BQ870873	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;					
REFERENCE	BQ870873	1 (bases 1 to 738)					
AUTHORS	BQ870873	Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L., and Bradford, K.					
TITLE	BQ870873	Lettuce and Sunflower ESTs from the Compositae Genome Project					
JOURNAL	BQ870873	http://compgenomics.ucdavis.edu/					
COMMENT	BQ870873	Unpublished (2002)					
	BQ870873	Contact: Alexander Kozik [R.W.Michelmore]					
	BQ870873	Department of Vegetable Crops, R.W.Michelmore Lab					
	BQ870873	University of California at Davis (UCD)					
	BQ870873	Asmundson Hall, UCD, Davis, CA 95616, USA					
	BQ870873	Tel: 1-(530)-742-1742					
	BQ870873	Fax: 1-(530)-752-9659					
	BQ870873	Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]					

belongs to contig OG\_CA\_Contig7026, see <http://cgpdb.ucdavis.edu/> for details.

Plate: OG10 row: E column: 14.

Location/Qualifiers

## FEATURES

source

```
1..738
/organism="Lactuca sativa"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone_lib="OG10E14"
/clone_lib="OG_ABCDI lettuce salinas"
/lab_host="E.coli"
/notes="Vector: pBRcDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformationally made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/ TAG_LIB=OG_ABCDI lettuce salinas TAG_TISSUE=roots environmental stress TAG_SEQ=ATCTCGCGG"
BASE COUNT      221 a 147 c 175 g 195 t
ORIGIN
Query Match      44.2%; Score 362.4; DB 14; Length 738;
Best Local Similarity 70.1%; Pred. No. 3.3e-96;
Matches 502; Conservative 0; Mismatches 211; Indels 3; Gaps 1;
Qy 13 ACTGCATTGCGAAGGCTAGAGGAAAGTTGCCCTTATAACAGGAGGAGCCAGTGGGA 72
Db 21 ACTAGTTCAACGAAGGCTAGAGGAAAGTTAGCATTCACCGGAGCAGCAAGCGGT 80
Qy 73 ATTGGAGAAGCACAGCAAAACTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGAT 132
Db 81 ATTSGTGAGTCACTGCAAAACTATTGTGGAACATGGAGCAAAAGTAGTCATTGCGAGAC 140
Qy 133 GTCAAGATGAATTAGTCTACTGCTGAGGCGCATTCGCATTCCTCAATGCCACTAC 192
Db 141 ATTCAGAACCAACTTGTCAAGCAGTTTGTGAGGCCATAGGCACATCAATCCATATAT 200
Qy 193 ATCCACTGTGATGTACTATGAGACGCTGTATAAATGCGGTGGCAACACAGTTTCA 252
Db 201 GTCCATTGTGATGAACCAATGAAGAGATGTCAAACGCTGTAGATCTGCAGTTTCC 260
Qy 253 ACCTATGAAACTGGACATTATGTTACAGATGCGAGGAATTTCTGATCCCAACAGGCC 312
Db 261 ACATACGAAAACTTGACATCATGTTTTCGAATGCGGGGATAGCAGATCTCTAAACGCCA 320
Qy 313 CGCATATAGACACGAAAGACAGACTTTGAAACGCGTTCTCAGTGTAAATGTAAACCGGA 372
Db 321 CGTATCGTTGATATGAAGACAGACACTTTGAGCGGTGTCTAGTGTCAATATACAGGT 380
Qy 373 GTTTTCCTATGATGAGCAGCAGCAGCTGTATGATTCAGCAGCAGCGAGTGGCAACATA 432
Db 381 GTCCTTTCTAAGCATGAACATGCTGTAGGGTTATGGTTTCCAGCAGCAGCTGCTCAATT 440
Qy 433 ATTTCCACTGCTAGTTTAAGCTCAACTATGGTGGTGGTGTCTTTCACATGCTTATTTGGT 492
Db 441 ATATCAACGGCTAGCTTGTGCTCAAAATATTGTTGGTGGCCTCGCATTTATTTGCT 500
Qy 493 TCAAGCATCTGTGTAGCCCTTACTAGCAATCTGGCAATCTGGCACTCGGACAAATTTGGC 552
Db 501 GCAAAACACGCTGTGCCCGTTTAACCAAAATCTTCAGTGGAACTTGGACAAATTTGGT 560
Qy 553 ATTAGGCTTAATGTTGCTCTCTTCGGCTTCTACGGCTTTTAGGCAAGAAATTTTCA 612
Db 561 ATAGAGTCAATTTGTTGCTCTCTTATGCAATGGTTCACACCACTAGCCCAAGTTTCGTT 620
Qy 613 CGGATTAATAATGAGAAATTTGAGATGTAATAAATCTTTGGCGGAAATTTGAAAGGT 672
Db 621 GGGCTTGA --- GGGGAACCAATTAGAAACTGTGTAACCTTACTTTGGGAATCTTTAAGGA 677
```

Qy 673 CCAAAATTATTTAGTGTGAGATGTTGCCAATGCAGCTCTTTATCTGCTAGTGATGA 728  
Db 678 CCACGCTCAGAACAGATGATGTTGCTAAGGCTGCACATTTTCTTGTGATGGCGA 733

## RESULT 2

BO875148

LOCUS

BO875148 704 bp mRNA linear EST 15-AUG-2002  
OG17E21.yg.ab1 OG\_ABCDI lettuce salinas Lactuca sativa cDNA clone  
DEFINITION  
BO875148 mRNA sequence.

ACCESSION

BO875148

VERSION

BO875148.1

KEYWORDS

EST

SOURCE

ORGANISM

Lactuca sativa

Lactuca sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; easterids II; Asterales; Lactuceae; Lactuceae;

Lactuca.

1 (bases 1 to 704)

REFERENCE

AUTHORS

Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,

Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison

,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,

Church,S., Jackson,L. and Bradford,K.

Lettuce and Sunflower ESTs from the Compositae Genome Project

<http://compgenomics.ucdavis.edu/>

Unpublished (2002)

Contact: Alexander Kozik [R.W.Michelmore]

Department of Vegetable Crops, R.W.Michelmore Lab

University of California at Davis (UCD)

Asmundson Hall, UCD, Davis, CA 95616, USA

Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659

Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]

belongs to contig OG\_CA\_Contig7026, see <http://cgpdb.ucdavis.edu/>

for details.

Plate: OG17. row: E column: 21.

FEATURES

source

1..704

/organism="Lactuca sativa"

/cultivar="Salinas"

/db\_xref="taxon:4236"

/clone\_lib="OG17E21"

/clone\_lib="OG\_ABCDI lettuce salinas"

/lab\_host="E.coli"

/notes="Vector: pBRcDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformationally made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/> TAG\_LIB=OG\_ABCDI lettuce salinas TAG\_TISSUE=roots environmental stress TAG\_SEQ=ATCTCGCGG"

BASE COUNT 215 a 139 c 165 g 185 t

ORIGIN

Query Match 42.5%; Score 348.4; DB 14; Length 704;

Best Local Similarity 70.3%; Pred. No. 4.5e-92;

Matches 482; Conservative 0; Mismatches 201; Indels 3; Gaps 1;

Qy 13 ACTGCATTCGCAAGGCTAGAGGAAAGTTGCCCTTATAACAGGAGGAGCCAGTGGGA 72

Db 21 ACTAGTTCAACGAAGGCTAGAGGAAAGTTAGCATTCACCGGAGCAGCAAGCGGT 80

Qy 73 ATTGGAGAAGCACAGCAAAACTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGAT 132

Db 81 ATTSGTGAGTCACTGCAAAACTATTGTGGAACATGGAGCAAAAGTAGTCATTGCGAGAC 140



```

Db 676 CTAACGCTCAGAACAGATGATGTTGC 701
RESULT 4
BE920368
LOCUS
DEFINITION BE920368 585 bp mRNA linear EST 02-OCT-2000
CSTB5H14 5' sequence, mRNA sequence.
ACCESSION BE920368
VERSION BE920368.1 GI:10446444
KEYWORDS
SOURCE EST:
potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE
1 (bases 1 to 585)
AUTHORS van der Hoeven,R.S., Bezzerides,J., Holt,I.E., Liang,F., Cho,J.,
UTterback,T., Hansen,C.L., Doan,B., Bougri,O., Buell,C.R., Ronning
,C.M., Fry,W.E., Tanksley,S.D. and Baker,B.
TITLE Generation of ESTs from potato leaves and petioles
JOURNAL Unpublished (2000)
COMMENT Contact: Cathy Ronning
The Institute for Genomic Research
For clone request: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com.
FEATURES
Source
loc_585
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="CSTB5H14"
/tissue_lib="potato leaves and petioles"
/tissue_type="leaflets and petioles"
/dev_stage="8 weeks old plants"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site1: EcoRI; Site2:
XhoI; Tissue was supplied by Dr. Fry (Cornell University).
Leaflets and petioles were isolated from 8 week old
greenhouse grown plants. The plants were watered and
fertilized freely. The tissue was immediately frozen in
liquid nitrogen."
BASE COUNT 164 a 154 c 139 g 128 t
ORIGIN
Query Match 35.5%; Score 290.6; DB 12; Length 585;
Best Local Similarity 71.6%; Pred. No. 5.8e-75;
Matches 396; Conservative 0; Mismatches 154; Indels 3; Gaps 1;
Qy 17 CATTCGCAAGAGGCTAGAGGAAAGTTGCCCTTATACAGAGGAGGCCAGTGGAAATTG 76
Db 33 CAATCGCAAGAGGCTAGAGGAAAGTAGCAATTGTACCGGAGGAGCTAGTGTATCG 92
Qy 77 GAGAAACACAGCAAACTCTTCCACATGGAGCCAAAGTTGCCATTGCTGATGTC 136
Db 93 GTGAAGCAATTGCAAGCTCTTCTGTGAACATGGAGCCAAAGTGTCAITCCGATGTC 152
Qy 137 AAGATGAATTAGTGTACTAGTGTTCGAGGCCAT--TGGCACTTCCAAATTCACCTACA 193
Db 153 AAGATGAACTCGGCAACTCAGTTCATCAACGCCCTCGGAGGCTCATCCAACTCCATTACA 212
Qy 194 TCCACTGTGATGTTACTAATGAGAGCGGTGTTAAATCCCGTGGACACACAGATTCAA 253
Db 213 TCCACTGTGAGTCACAAGCAATACACGCTCCAAGAAAGCTGTGCACAGAACCATCGCA 272
Qy 254 CCTATGAAACCTGGACATTATGTTCCAGCAATGCAGGAATTTCTGATCCCAACAGGCC 313
Db 273 CATTTGAAACCTCGACATCATGTCATGCAATGCTGGCATATGGCAGACCAAGTCAA 332
Qy 314 GCATCATGACACAGCAAGAGACTTTGAACGGTCTCAGTGAATGTAACCGGAG 373
Db 333 GATTCATAGACACACAGACAGATTTTCGAACGAGTCTTAGCATCAACGTGACAGGAG 392

```

---

```

Qy 374 TTTTCTATGCATGAAGCAGCGACGCTGTTATGATTCAGCAGCAGCTGGCAACATAA 433
Db 393 TTTTCTTGAGCATGAAGCATCGCGCTCATGCTGCTCCATGCGTAGCGCTGCATCA 452
Qy 434 TTTCCACTGCTAGTGTAAAGCTCAACTATGATGTTGTTCTTTCACATGCTATTGTTG 493
Db 453 TCTCCACCGCTAGTGTAAAGCTCCACGCTTGGAGCTGCAGCATCCCATGCTCAGCT 512
Qy 494 CAAGCATGCTGTGTTAGCCCTTACTAGGAATCTGCAGCTCGAGCTCGACAAATTGGCA 553
Db 513 CCAACACGCGCTGTGTTGGCTCACCAGAACCTGAGCTGGAGCTAGGACAAATTGGTA 572
Qy 554 TTAGGTTAATTG 566
Db 573 TACGTGTGAATTG 585
RESULT 5
BG350846
LOCUS
DEFINITION BG350846 676 bp mRNA linear EST 01-MAR-2001
101C04 Mature tuber lambda ZAP Solanum tuberosum cDNA, mRNA
sequence.
ACCESSION BG350846
VERSION BG350846.1 GI:13179588
KEYWORDS EST
SOURCE Potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE
1 (bases 1 to 676)
AUTHORS Nielsen,K.L., Crookshanks,M., Emmersen,J. and Welinder,K.G.
TITLE EST-sequencing of mature potato tuber (Var. Kuras)
JOURNAL Unpublished (2000)
COMMENT Contact: Karen G. Welinder
Institut for bioteknologi
Aalborg Universitet
Sølngeaardsholmsvej 49, 9000 Aalborg, Denmark
Tel: +45 96358467
Fax: +45 98141808
Email: kgw@bio.auc.dk
Sequenced from the 5' end.
High quality sequence stop: 676
POLYA=NO.
FEATURES
Location/Qualifiers
1..676
/organism="Solanum tuberosum"
/cultivar="Field grown Kuras"
/db_xref="taxon:4113"
/clone_lib="Mature tuber lambda ZAP"
/tissue_type="tuber"
/note="Vector: Lambda ZAP"
BASE COUNT 187 a 176 c 157 g 156 t
ORIGIN
Query Match 35.2%; Score 288.4; DB 12; Length 676;
Best Local Similarity 70.4%; Pred. No. 2.8e-74;
Matches 401; Conservative 0; Mismatches 166; Indels 3; Gaps 1;
Qy 18 ATTCGCAAGAGGCTAGAGGAAAGTTGCCCTTATAACAGGAGGAGCGAGTGGAAATTGG 77
Db 88 ATCGGCAAGAGGTTAAAGGGGAAAGTAGCAATGTTAACCGGAGGAGCTAGTGGTATCG 147
Qy 78 AGAAACACACAGCAAACTCTTCTCCCAACATGAGGCAAAAGTTGCCATTGCTGATGTCCA 137
Db 148 TGAAGCAATTGCAAAAGCTCTTCTGCAACATGAGGCAAAAGTGGTCAITTCGATGTCCA 207
Qy 138 AGATGAATTAGTGTACACTAGTGTTCGAGGCCAT--TGGCACTTCCAAATTCACCTACAT 194
Db 208 AGATGAACCTCGGCAACTCAGTTCATCAACACCCCTCGGAGGTTTCATCAATTCATTTAG 267
Qy 195 CCACTGTGATGTTACTAATGAAGACGCTGTTAAATGCGGTGGGACACACAGTGTCAAC 254

```



Db	268	CCACTCGGACGTCACAAACGAACACCGTCCAAAGAGCTGTCGACAGAACCAATCCGCCAC	327
Qy	255	CTATGGAAAACTGGACATTTATGTTTCAGCAATGCAGGAATTTCTGATCCCAACAGGCGCG	314
Db	328	ATTTGGAAAACTCGACATCATGATCTGCAATGCTGGCATATGCGACGAGACCAAGTCAAG	387
Qy	315	CATCATAGACAACGAAACAGACATTTGAAACGGTTCTCAGTGTAAATGTAAACCGGAGT	374
Db	388	AATCATAGACAACGAAACGAGATTTTCGAAACGAGTCTTTAGCATCAACGTAACAGGAGT	447
Qy	375	TTTCTCTATGATGAAGCACGACGACGTCGTTATGATTTCCAGCAGCAGTGGCAACATAAT	434
Db	448	TTTCTTGAGCATGAAGCATCCGCGTCATGGTCCCGACGCGTAGCGGCTGCATCAT	507
Qy	435	TTCCACATGCTAGTTTAAAGTCACATATGGGTGGGTCTTTCACATGCCTATTGTGCGTTC	494
Db	508	CTCCACCGCTAGTGTAGCTTCCAGGTCGGAGCTGCAGCATCCCATGTCATCTGCAGCTC	567
Qy	495	AAAGCATGCTGTGTTAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGACAATTTGGCAT	554
Db	568	CAACACGCTGTGTTGGGGTCCACCAAGACCTAGCGGTGGACCTAGGACAATTCGGTAT	627
Qy	555	TAGGGTAAATGTTTGTCTCCTTTCCGGCT	584
Db	628	ACGTGTCAATTTGTTGTACCAATACCGAT	657

RESULT 6  
BM812901  
LOCUS  
DEFINITION  
EST590994 BNIR Medicago truncatula/Meloidogyne incognita mixed EST  
library cDNA clone pBNIR-4O10, mRNA sequence.  
ACCESSION  
BM812901  
VERSION  
BM812901.1 GI:19148915  
KEYWORDS  
EST.  
SOURCE  
Medicago truncatula/Meloidogyne incognita mixed EST library.  
ORGANISM  
Medicago truncatula/Meloidogyne incognita mixed EST library  
Eukaryota: mixed EST libraries.  
REFERENCE  
1 (bases 1 to 717)  
AUTHORS  
Bird,D., Koital,H., Samac,D., Town,C.D., Van Aken,S., Utterback,T.,  
Cheung,F., Tsai,J. and Fraser,C.M.  
TITLE  
ESTs from roots of Medicago truncatula after infection with the  
nematode Meloidogyne incognita  
JOURNAL  
Unpublished (2002)  
COMMENT  
Contact: Bird, DM  
Plant Nematode Genetics Group  
North Carolina State University  
Box 7616, Raleigh, NC 27695, USA  
Tel: 919 515 6813  
Fax: 919 515 9500  
Email: david\_bird@ncsu.edu  
TIGR sequence name: MTQAB89TK More information is available at:  
www.medicago.org  
Seq primer: SKmod (CTA gAA CTA qTq qAT CC).

DB	B50	GCTGAAGCTGCTTTGTATTTCGGGAAGTGATGAGTCTAAGTATGTTAGTGGTGCTTTAAATCTT	709
QY	757	TTCATTGA 764       	
DB	710	GTCATAGA 717	
RESULT	7		
LOCUS	BQ630495		
DEFINITION	BQ630495 637 bp mRNA linear EST 02-JUL-2002 sq08rb06.y1 Gm-cl045 Glycine max cDNA clone SOYBEAN CLONE ID: 5', similar to TR:P93697 P93697 CPD12 PROTEIN. ;, mRNA sequence.		
ACCESSION	BQ630495		
VERSION	BQ630495.1	GI:21678144	
KEYWORDS	EST		
SOURCE	soybean.		
ORGANISM	Glycine max		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Rosales; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;		

RESULT 8					
BE205431					
LOCUS	BE205431	686 bp	mRNA	linear	EST 05-SEP-2000
DEFINITION	EST398107	KV0 Medicago truncatula	cDNA clone pKV0-21J9.		mRNA

ACCESSION	BE205431.1	GI:8748729
VERSION		
KEYWORDS	barrel medic.	
SOURCE	Medicago truncatula	
ORGANISM	eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabiales; Fabaceae; Papilionoideae; Trifoliales; Medicago.	
REFERENCE	1. (bases 1 to 686)	
AUTHORS	Vandenbosch K., Endre, G., Hur, J., Moore, J., Beremand, P., Ellis, L., Town, C.D., Bowman, C.L., Craven, W.B., Hansen, T.S., Holt, I.E. and Fraser, C. M.	
TITLE	ESTs from unincubated seedling roots of Medicago truncatula	
JOURNAL	Unpublished (1999)	
COMMENT	Contact: Vandenbosch K Department of Plant Biology University of Minnesota 220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA Tel: 612 624 2755 Fax: 612 625 1738	

```

TEXAS A&M UNIVERSITY NAME:12852036
TIGR sequence name:TXGBM53TK
More information is available at:
http://chrystie.tamu.edu/medicago
Seq primer: SKmod (CTA GAA CTA gtg gat CC).
Location/Qualifiers
1. .686
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pkV0-21J9"
/clone_lib="KV0"
/tissue_type="Seedling roots"
/dev_stage="Immediately prior to inoculation with
Sinorhizobium meliloti (0 hour)"
/lab_host="E.coli strain XL0LR"
/note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in XL0LR cells."
197 a 126 c 166 g 197 t
BASE COUNT
ORIGIN
Query Match 31.6% Score 258.4 DB 10 Length 686

```





JOURNAL COMMENT	Unpublished (2000) Contact: Deborah A. Samac Department of Plant Pathology University of Minnesota 495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA Tel: 612 625 1243 Fax: 651 649 5058 Email: debbys@puccini.crl.umn.edu TIGR sequence name: M278844e More information is available at: http://chrysie.tamu.edu/medicago Seq primer: SKmod (CTA gAA CTA gTg gAT CC).	
FEATURES	Location/Qualifiers 1..681 /organism="Medicago truncatula" /cultivar="genotype AL7" /db_xref="taxon:3880" /clone="pDSIL-43C22" /tissue_type="leaves infected with Colletotrichum trifolii" /dev_stage="cotyledons and primary leaves harvested 5 and 8 days after inoculation with Colletotrichum trifolii" /lab_host="E. coli strain XL0R" /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from cotyledons and primary leaves harvested 5 and 8 days after inoculation with Colletotrichum trifolii. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XL0LR cells. Note: EST may be of fungal origin."	
source	195 a 126 c 168 g 192 t	
BASE COUNT		
ORIGIN		
Query Match	31.4%;	Score 257.4; DB 12; Length 681;
Best Local Similarity	62.0%;	Pred. No. 4.4e-65;
Matches	430; Conservative	0; Mismatches 251; Indels 12; Gaps 1;
QY	52	ATAACAGGAGGCCGATGGAATGGAGAACACACAGCAAACTCTTCCCAACATGGA 111
Db	1	ATCACCAGGTGGTCGAGCGGCATGTGTAGGCCACTGCAAGACTCTTCCGAGCATGGA 60
QY	112	GCCAAAGTTGCCATGCTGATGCCAAGATGAATTAGGTCACATCAGTGTGTCGAGGCCATT 171
Db	61	GCTCAAGTCGTTATAGCTGACATCCAGACGACAAGGACACTCTATTTCGAAGGAATTA 120
QY	172	GGCACTTCCAAATCCACCTACATCCACTGTGATGTTTACTAATGAAGACGGTGTAAAAAT 231
Db	121	CAAAAATCGTCTTCTTCCATGTTCTGTTGACGTGACAAAGGAAGACATTTGAAAT 180
QY	232	GCGGTGGACACACAGTTTCAACCTATGAAAACTGGACATATATGTCAGCAATGCAGGA 291
Db	181	GCGGTGAACACACACCGTTTCAAGTACGGTAACTAGATATCATCTTTAAACACGCTGGC 240
QY	292	ATTCTTGATCCCAACAGGCCCCCATCATAGACAACGAAAGACACATTTGAACCGCTT 351
Db	241	ATAAGCGGTGTAAACAACAACAAGATACTTTGAAACAATATCTGAATTTGAGGATGTG 300
QY	352	CTCAGTGTAAATGTAACCGGAGTTTTCCTATGCATGAAGCAGCAGCAGCTGTGTATGATT 411
Db	301	ATTAAAGTAAACTTGACTGCGCTCTTCTTGGAAACAACAATGCACAGAGGGTAATGATC 360
QY	412	CCAGCAGCAGTGGGCAACATAATTCACCTGCTAGTTTAAGCTCACTATGGGTGGTGTG 471
Db	361	CTGTCTCGACGGGTAGCATAATTAACACAGCTAGTGTGGTGGGAAGCATTTGGTGGTGT 420
QY	472	TCTTCATCGCTTATGTGTGGTCAAGACATGCTGTGTTAGCCCTTACTAGGATCTGGCA 531
Db	421	GTCCACATGCCTTACACAAGTTCAAGACGCGCGTGTGTTGGGCTGATGAGAAACACGCT 480
QY	532	GTGAGCTCGGACAATTTGGCATTAGGTTAATGTTTGTCTCTTTCGGGCTTCCTACG 591
Db	481	ATCGAGCTTGGGCAATTCGGTATTCGGTGAAATGTGTGTCGCTTATTTCTTGGCACT 540
QY	592	GCTTTAGGCAAGAAATTTTCAGGGATTAAAAATGAAGAAGAATTTTGAGATGTAAATAAC 651
Db	541	CCGATGTTTACAAATTTCTTTAAGCTTGATGATGGTGGAGCTCTAGATATTTTTTCTAAC 600
QY	652	TTTGGCGGAAATTTGAAAGGTCACAAATTTAATGTTTGAGGATGTTGCCAATGCAGCTCTT 711
Db	601	CTAAAGGGTACTAATCTTGTGCCAAAA-----GATGTGCTGAAGCTGCTTTG 648
QY	712	TATCTGGCTAGTGATGAGGCAAAATACGTGAGT 744
Db	649	TATTTGGGAAGTGATGAGTCTAAGTATGTAGT 681
RESULT 12		
BQ862029		
LOCUS	BQ862029	695 bp mRNA linear EST 14-AUG-2002
DEFINITION	OGClh09_yq_ab1 OG_ABCDI lettuce salinas Lactuca sativa cDNA clone	
ACCESSION	BQ862029	OGClh09 mRNA sequence.
VERSION	BQ862029.1	GI:22247494
KEYWORDS	EST.	
SOURCE	Lactuca sativa.	
ORGANISM	Lactuca sativa.	
REFERENCE	1 (bases 1 to 695)	
AUTHORS	Kozik,A., Michelmoe,R.W., Knapp,S., Matvienko,M., Rieseberg,L., Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z., Church,S., Jackson,L. and Bradford,K.	
TITLE	Lettuce and Sunflower ESTs from the Compositae Genome Project	
JOURNAL	Unpublished (2002)	
COMMENT	Contact: Alexander Kozik [R.W.Michelmoe] Department of Vegetable Crops, R.W.Michelmoe Lab University of California at Davis (UCD) Asmundson Hall, UCD, Davis, CA 95616, USA Tel: 1-(530)-742-1742 Fax: 1-(530)-752-9659 Email: akozik@ucdavis.edu [michelmoe@vegmail.ucdavis.edu] singleton, see http://cgdb.ucdavis.edu/ for details. Plate: QGCl row: h column: 09.	
FEATURES	Location/Qualifiers	
source	1..695 /organism="Lactuca sativa" /cultivar="Salinas" /db_xref="taxon:4236" /clone="OGClh09" /lab_host="E.coli" /note="Vector: pBRCDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgdb.ucdavis.edu/TAG_LIB-OG-ABCDI lettuce salinas TAG_TISSUE=chemical TAG_SEQ=GTGAGCCGG"	
BASE COUNT	208 a 136 c 163 g 186 t	2 others
ORIGIN		
Query Match	31.3%;	Score 256.6; DB 14; Length 695;
Best Local Similarity	61.4%;	Pred. No. 7.6e-65;



```

ACCESSION   BQ982959
VERSION     BQ982959.1  GI:22400484
KEYWORDS    EST.
SOURCE      Lactuca sativa.
ORGANISM    Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.
REFERENCE   1 (bases 1 to 544)
AUTHORS    Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
            ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
            Church,S., Jackson,L. and Bradford,K.
TITLE      Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compenomics.ucdavis.edu/
JOURNAL     Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA.Contig3351, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGE17 row: N column: 08.
            Location/Qualifiers
                1..544
                /organism="Lactuca sativa"
                /cultivar="L.serriola"
                /db_xref="taxon:4236"
                /clone="QGE17N08"
                /clone.lib="QG.EFGHJ lettuce serriola"
                /lab_host="E.coli"
                /note="Vector: pBRCDNASFIAB; The library was constructed
                from 10 different sources of RNA from a single genotype.
                Separate cDNAs were generated using primers that
                incorporated unique 5' and 3' tags to distinguish each
                source of RNA. cDNAs were then pooled, size-fractionated,
                directionally cloned into a custom medium-copy vector and
                transformations made with four size classes to minimize
                size bias. Details of each source of RNA and library
                construction can be obtained at http://cgpdb.ucdavis.edu/
                TAG_LIB-QG.EFGHJ lettuce serriola
                TAG_TISSUE=leaves dark grow
                TAG_SEQ=GCTAGTCGGG"
BASE COUNT  159 a 95 c 123 g 167 t
ORIGIN
Query Match 30.8%; Score 252.6; DB 14; Length 544;
Best Local Similarity 67.6%; Pred. No. 1e-63;
Matches 370; Conservative 0; Mismatches 174; Indels 3; Gaps 1;

QY 259 GGAAACTGGACATTATGTTACGAAATTCGATCCCAACAGCCCGCAATC 318
Db 1 GGAATACTTGATATCATGTTTGTATGAGGAATAATCGCTTAACAAGGATCACATC 60

QY 319 ATAGCAACGAAAGACACACTTGAACCGGTTCTCAGTGTAAATGTACCGGAGTTTC 378
Db 61 ATCGATAATGAAATAATTCGATTGTTGAACGTGGTGTAGTGTCAACGTCACAGGTGTCCTTT 120

QY 379 CTATGATCAAGCACGACGCTGTATGATTCCAGACGACGAGTGGCAACATAATTTCC 438
Db 121 CTAGCATGAACATGACGCTAGGTTATGTTTCCACACAGAGCTGGGTCAATATATCA 180

QY 439 ACTGCTAGTTTAAAGCTCAACTAGTGGTGGTCTTCTACATGCCTATTGTGGTTCAAG 498
Db 181 ACAGCTAGTGTGTGCAAAATATCGTGGCTGTGCCCCACATGCTTACACTGTTTCAAG 240

QY 499 CATGCTGTGTTAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGACAAATTTGGCAT 558
Db 241 CATGCTGTAGCTGGTTTAAACCAAGAATCTTGCAGTTGAGCTTGGACAAATTTGTTTCA 300

```

---

```

QY 559 GTTAATTTGTTCTCTCTTTTCGGGCTTCCTACGGCTTTTAGCAAGAAATTTTCAGGGATT 618
Db 301 GTCAATTGTCCTTTACCTTTACGCAATAGATACACCACAGCCACAGGTTTCACGGGTCTT 360

QY 619 AAAATGAAGAAGAAATTCGAGAAATGAATAAACTTTTCGGGAAATTTGAAGGTCCAAA 678
Db 361 AAA---GGGGAGGATTTCAAGAACAGAGGTAACACTCAGTTTGCAACCTTAAGGCTGTGACA 417

QY 679 TTAAATGTTGAGGATGTTCCCAATGCAGCTCTTTATCTCGCTAGTCAGCAAAATAC 738
Db 418 CTTACAACAGATGATGTTCTTAAGCTGCTGTTTTTTTGGCGAGTGATGAGCTAAGTAT 477

QY 739 GTGAGTGGACAAATCTGTTTCATTATGATGAGGTTCCAGGCTCTGCAATTTCTGTAAATCAA 798
Db 478 ATTAGTGGACAAATTTGTTTATTGATGTTGATTTAGCATTTGGTAACCCATCATTTCAAT 537

QY 799 GTGTTCC 805
Db 538 ATGTTTC 544

RESULT 15
BQ914508
LOCUS     BQ914508
DEFINITION BQ914508.1 715 bp mRNA linear EST 19-AUG-2002
            clone QHB10M20, mRNA sequence.
ACCESSION BQ914508
VERSION   BQ914508.1  GI:22313289
KEYWORDS  EST.
SOURCE    common sunflower.
ORGANISM  Helianthus annuus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
            1 (bases 1 to 715)
            Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
            ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
            Church,S., Jackson,L. and Bradford,K.
            Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compenomics.ucdavis.edu/
            Unpublished (2002)
            Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            singleton, see http://cgpdb.ucdavis.edu/ for details.
            Plate: QHB10 row: M column: 20.
            Location/Qualifiers
                1..715
                /organism="Helianthus annuus"
                /cultivar="RHA801"
                /db_xref="taxon:4232"
                /clone="QHB10M20"
                /lab_host="E.coli"
                /note="Vector: pBRCDNASFIAB; The library was constructed
                from 11 different sources of RNA from a single genotype.
                Separate cDNAs were generated using primers that
                incorporated unique 5' and 3' tags to distinguish each
                source of RNA. cDNAs were then pooled, size-fractionated,
                directionally cloned into a custom medium-copy vector and
                transformations made with four size classes to minimize
                size bias. Details of each source of RNA and library
                construction can be obtained at http://cgpdb.ucdavis.edu/
                TAG_LIB-QH_ABCDI sunflower RHA801
                TAG_TISSUE=callus
                TAG_SEQ=GTACTGCGG"
FEATURES
source

```

BASE COUNT		218 a	140 c	161 g	196 t		
ORIGIN							
Query Match		30.8%; Score 252.6; DB 14; Length 715;					
Best Local Similarity		65.9%; Pred. No. 1.2e-63;					
Matches 366; Conservative		0; Mismatches 189; Indels 0; Gaps 0;					
QY	23	CAAGAAGCTAGAGGAAAGTTGCCCTTATAACAGAGGAGGCCAGTGGAAATTTGGAGAAA	82				
DB	94	CACCAAGATTGGCAGGTAGGTAGCTTTTATTTTATTTTATTTTATTTTATTTTATTTT	153				
QY	83	CCACAGCAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGTGATGTCCTCAAGATG	142				
DB	154	CAACTGCARAACTATTGCTTAACATGGGGCGGAAAGTTATTATTGCCGATATCCAAAGATG	213				
QY	143	AATTAGTCACTCAGTTGTCGAGGCCATTGGCACTTCCAAATTCACCTACATCCACTGTG	202				
DB	214	ATTGGGTCAATCTGTTGTGAAGAAATAGGCCCTTGAAATGCATCTTATGTCCACTGTG	273				
QY	203	ATGTTACTAATGAAGCGGTGTTAAAAATGCCGTGGACACACAGTTTCAACCTTATGCAA	262				
DB	274	ATGTAACCTGTTGAGGCTGATGTAGAAATGTTGTAACCTTTGCCACACAAAGTATGGTA	333				
QY	263	ACTGGACATTATGTCAGCAATTCAGGAATTTCTGATCCCAACAGGCCCGGCATCATAG	322				
DB	334	ACTAGATATTATGATCAACAATCGCGCTATTCTAGGTCAAGCGAACTCAGCATTTCTG	393				
QY	323	ACACGAAAGACAGACTTTGACCGGTTCTCAGTGTAAATGTAACCGGAGTTTTCCTAT	382				
DB	394	ACACGCAAAAGCTGATTTGATCGTTGTAAACGTGAATCTAGCGGCGGTTCCTTG	453				
QY	383	GCATGAAGCAGCGCAGCGTGTATGATTTCCAGCAGCGGATGGCAACATAATTTCCACTG	442				
DB	454	GAATGAAGCAGCGTGTCTGTCGCATGATTTCCAGCGCTTGTGGTAGCATATCTCGATTG	513				
QY	443	CTAGTTTAAAGCTCAACTATGGTGGTGGTCTTTCACATGCCTATTGTGGTTCAAAGCATG	502				
DB	514	GGAGTGTCTTCTGCTATCGGCGGTGTTTCCCCACACGCGTATACGAGTACCAAAACATG	573				
QY	503	CTGTGTTAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGACAATTTGGCATTAGGGTTA	562				
DB	574	CAATTTTGGACTCAGCAAGAATGTGGCGCAGAGCTTGGCAAAACATGGAATTCGTGTCA	633				
QY	563	ATTGTTTCTCTCCTT	577				
DB	634	AATGTGTATCACCAT	648				

Search completed: February 18, 2003, 17:45:38  
Job time : 1475 secs



Result No.	Score			Query		DB	ID	Description
	Score	Match	Length					
1	114.8	14.0	1187	1	US-08-440-856A-2			Sequence 2, Appli
2	100	12.2	1288	1	US-08-440-856A-9			Sequence 9, Appli
3	97.2	11.9	1236	1	US-08-440-856A-1			Sequence 1, Appli
4	70.2	8.6	771	4	US-09-504-358-13			Sequence 13, Appl
5	70.2	8.6	771	4	US-09-954-314-13			Sequence 13, Appl
6	70.2	8.6	10629	4	US-09-504-358-15			Sequence 15, Appl
7	70.2	8.6	10629	4	US-09-954-314-15			Sequence 15, Appl
8	60.6	7.4	732	4	US-09-239-052-1			Sequence 1, Appli
c	59.6	7.3	19702	4	US-08-961-527-7			Sequence 7, Appli
10	57.8	7.1	11196	4	US-08-859-207-8			Sequence 8, Appli
11	54.4	6.6	873	4	US-09-134-001C-1502			Sequence 1502, Ap
12	54	6.6	774	4	US-09-134-001C-725			Sequence 725, App
13	52.8	6.4	2774	4	US-09-363-189B-5			Sequence 5, Appli
14	51.2	6.3	825	4	US-09-134-001C-1594			Sequence 1594, Ap
15	50.4	6.2	786	4	US-09-468-738A-28			Sequence 28, Appl
16	50.4	6.2	786	4	US-09-940-019-28			Sequence 38, Appl
17	49.2	6.0	750	4	US-09-134-001C-1988			Sequence 1988, Ap
c	45.4	5.5	4403765	4	US-09-103-840A-2			Sequence 2, Appli
19	43.6	5.3	804	4	US-09-134-001C-2205			Sequence 2205, Ap
20	42.6	5.2	1995	4	US-08-936-165A-250			Sequence 250, App
21	41.6	5.1	852	4	US-09-134-001C-2509			Sequence 2509, Ap
22	40	4.9	756	3	US-08-822-322-7			Sequence 7, Appli
23	40	4.9	756	4	US-09-466-109-7			Sequence 7, Appli
24	39.6	4.8	650	4	US-09-468-738A-13			Sequence 13, Appl
25	39.6	4.8	650	4	US-09-940-019-13			Sequence 13, Appl
26	39.6	4.8	879	4	US-09-468-738A-1			Sequence 1, Appli
27	39.6	4.8	879	4	US-09-940-019-1			Sequence 1, Appli

Db 213 ATCGGGAGCGCATGTGAGGCTGTTTAAACACGGGGCAAGGTGGTATCGCGGAC 272  
Qy 133 GTCCAGATGAATAGGTCACTCAGTTCGAGGCGATTGGCACTTCCAATTCACACTAC 192  
Db 273 ATCGACACGCGGCGGAGGCGCTGGCGGGCGCTGGG--GCCGACGTCGGGTTTC 329  
Qy 193 ATCCACGTGATGTACTAATGAAGACGGTGTAAATGCGGTGGACACACAGTTTCA 252  
Db 330 GTGCGGTGCCAGCTGTGCTGGAGGAGGACGTGGAGCGCGCGCTCGAGCGCGCGTGGCG 399  
Qy 253 ACCTATGGAACACTGACATATGTTAGCAATGACAGGAATTTCTGATPCCCAACAGGCC 312  
Db 390 CGGTACGCGCGCGTGCAGCTGTGTGCACACACGCGCGGGGTCTGGCGCCGACAGCGCG 449  
Qy 313 -----CGCATATAGACACACAAAGACAGACTTTGAACGCGTTCCTCAGTGAAT 363  
Db 450 GCCGCCAAGASCATCTCGTTCGACGCGGGGAGTTTCGACCGCGTCTCGCGTCAAC 509  
Qy 364 GTACCCGAGCTTTTCTATGCAATGAAGACGACGACGCTGTATGATTCAGACACGAGT 423  
Db 510 GCGCTGGCGCGCGCTGCGCATGAAGACGCGCGCTCGCCATGACCCAGCGCGCGCC 569  
Qy 424 GGCAACATAATTCACACTGCTAGTTTAAAGCTCAACTATGCGTGTGTTCTTCACATGCC 483  
Db 570 GCGACATCATCTCGCTCGCAGCGCTCGCGCGCGTCTCGCGCGCTCGCGCGCGCGCC 629  
Qy 484 TATTGTGTTCAAGCATCTGCTGTAGCCCTTACTAGGAATCTGGCAGTCTCGAGCTCGGA 543  
Db 630 TACACCGCTTCCAGCACCGCCATCTGGGGCTCACCAGAACCGCGCTCGAGCTCGGC 689  
Qy 544 CAATTTGGCAATAGGTTTAATGTTGTCCTTTTCGGGGCTTCTAGCGCTTTAGGCAAG 603  
Db 690 GCCACGCGCATCGCGTCAACTGCACTCCCGCTTCGCGCGTGGCCACCGCGATGCTCATC 749  
Qy 604 AA 605  
Db 750 AA 751

## RESULT 2

US-08-440-856A-9  
; Sequence 9, Application US/08440856A  
; Patent No. 5750873  
; GENERAL INFORMATION:  
; APPLICANT: DELLAPORTA, STEPHEN L.  
; TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING  
; TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 PENNSYLVANIA AVE. N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/440.856A  
; FILING DATE: 15-MAY-1995  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MILLMAN, ROBERT A.  
; REGISTRATION NUMBER: 36,217  
; REFERENCE/DOCKET NUMBER: 05463-20001.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1517  
; TELEFAX: (202) 887-0763  
; TELEX: 706141

; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1288 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-08-440-856A-9  
  
Query Match 12.2%; Score 100; DB 1; Length 1288;  
Best Local Similarity 47.9%; Pred. No. 4.3e-22;  
Matches 292; Conservative 39; Mismatches 259; Indels 20; Gaps 4;  
  
Qy 13 ACTGCATTCGAAGAAGGCTAGAAGGAAAGTTGCCCTTATAACAGGAGGAGCGAGTGA 72  
Db 231 ACWCASATGCCAAGAGGCTGGAMGGAAAGTGGCCATTGTSAAGCGGCGGCGGSGGG 290  
Qy 73 ATTGGAGAAACACACAAACACTTCTCCCAACTGAGGCAAAAGTTGCCATTTGCTGAT 132  
Db 291 ATCGGAGAGCSATCTGTMGGCTGTTTCGYAAGACACGGCGCCMRGGTGTATCCGCGAC 350  
Qy 133 GTCCAAGATGAATTAAGTTCACATGCTGTCAGGAGGCGCATTTGGCACTTCCCAATTCACCTAC 192  
Db 351 ATCGACACCGCGCGGSGAGGCGCTGGCGKCGCGCTGGG---CCCAAGTCRGSFTC 407  
Qy 193 ATCCACTGTGATGTTACTAATGAAGACGCTGTAAATGCGGTGGAGACACACAGTTTCA 252  
Db 408 GTGCGSTGCCAGCGTGTCTGTGGAGGAGACGTSRGGCGCGCGTSGACTGGCGCTGSG 467  
Qy 253 ACCTAT---GGAAACTGGACATTAATGTCACAAATGACAGGAATTTCTGATPCCCAACAG 309  
Db 468 CGSRYGRGCGCGCTSGACGTSYSTGCAACACCGCGGGGTGCTGGGCGCGCGACAGC 527  
Qy 310 CCC-----CGCATATAGACACACGAAAGACAGACTTTGAACGCGTTCTCAGTGA 360  
Db 528 CGCGCGCGCCARGAGCATCTGCTTCGACGCGSSGAGTTTCGACCGCGTCTCGCGCTC 587  
Qy 361 AATGTAACCGAGTTTTCCTATGCAATGAAGACGACGACGACGCTTATGA-----TTCAG 415  
Db 588 AAGCGCTGGCGCGCGCTCGSATGAAGACACGCGGCGCKCGCATGACCCAGCGCGCC 647  
Qy 416 CAGCAGTGGCAACATAATTTCCACTGCTAGTTTAAAGCTCAACTATGGTGGTGGTTCTT 475  
Db 648 GCCGCGGGGAGCATCTCTCGCTCGCAGCGCTCGCSGCGCTGCTSGGCGGCTCGGCC 707  
Qy 476 CACATGCTATTGCTGTTTCAAGCATGCTGTGTAGCCCTTACTAGGAATCTGCAGTCG 535  
Db 708 CGACGCTACACCGCTTCAAGCAGCCATCGTSGGCTCACCAGAACCGCGCTGCG 767  
Qy 536 AGCTCGGCAATTTGGCATTAGGTTAATTTGTTCTCTCTTTCGGGCTTCTCAGGCTT 595  
Db 768 AGCTSSGCGSCACGGSRTCCGSGTCAACTGCTCTCCCTTCGGGCTGCCACGCCSA 827  
Qy 596 TAGGCAAGAA 605  
Db 828 TGCTCATCAA 837  
  
RESULT 3  
US-08-440-856A-1  
; Sequence 1, Application US/08440856A  
; Patent No. 5750873  
; GENERAL INFORMATION:  
; APPLICANT: DELLAPORTA, STEPHEN L.  
; TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING  
; TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 PENNSYLVANIA AVE. N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA

ZIP: 20037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release 1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 15-MAY-1995  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: WILLIAM ROBERT A.  
REGISTRATION NUMBER: 36,217  
REFERENCE/DOCKET NUMBER: 05463-20001.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1517  
TELEFAX: (202) 887-0763  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1236 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-440-856A-1

Query Match 11.9%; Score 97.2; DB 1; Length 1236;  
Best Local Similarity 51.1%; Pred. No. 3.3e-21;  
Matches 312; Conservative 0; Mismatches 283; Indels 15; Gaps 3;

QY 8 TTCGAACCTCATTCCGAAGAAGCTTAGAAGAAAGTTGCCCTTTATAACAGGAGGACCA 67  
DB 220 TGCACACACCGATGCCAAGAGCTGGACGGAAGGTGGCCATTGTGACGGCGGGCGGC 279  
QY 68 GTGGAAATTCGAGAAACACAGCAAACTCTCTCCCAACATGAGCCAAAGTTGCCATTG 127  
DB 280 GCGGGATCGGAGGCCATCGTGGCGCTGTCGCCAAGCAGCGGGCGGGGGTGGTATCG 339  
QY 128 CTGATGTCGAAGATGAATTAAGTCACTAGTTCGAGGCCATTGGCACTTCCAAATCCA 187  
DB 340 CGGACATCAGCAGCGCGGGGAGGCGCTGCGTFCGCGCTGGGCCGC ---AGGTCA 396  
QY 188 CCTACATCAGTGTATGTTACTAATGAAGACGGTGTAAATGCGCGTGGACACACAG 247  
DB 397 GCTTCGTGCGCTGCGACGTGTCGTTGGAGAGCAGCTCCGCGCGCGCGTGGACTGGCGC 456  
QY 248 TTTCAACCTA---TGGAAACTGGACATTATGTCAGCAATCAGAAATTTCTGATCCCA 304  
DB 457 TGTGCGGCCAGCGCGCGCGCTCGAGCTTACTGCAACCAACCGCGGGGTGCTGGCGCGCC 516  
QY 305 ACAGGCC-----CGCATCATAGACAAGCAAGCAAAAGCAGACTTTGAACGCTTCTCA 355  
DB 517 AGACGCGCGCGCCAGGACATCTGCTCTCGACGCGCGCGAGTTCGACGCGGTCTCC 576  
QY 356 GTGTAATGTACCGAGTTTTCCTATGATGATGAAGCAGCAGCAGCGTGTATGATTCAG 415  
DB 577 GGTCAACCGCTGGCGCGCGCTCGGATGAAGCAGCAGCGCGCGCCATGGCGCGCC 636  
QY 416 CAGCAGTGGCAACATAATTTCCACTGCTAGTTTAGCTCAACTATGGGTGGTGTCTTT 475  
DB 637 GCGCGCGGGAGCATGCTCTCCGTGCGCAGCTGCGCGCGCTGCTGGCGCGCTCGGCC 696  
QY 476 CACATGCTATTGTGGTTCAAGACATGCTGTGTAGCCCTTACTAGGAATCTGGCAGTGC 535  
DB 697 CGCAGCCTACACCGCTCCAGACCGCCATCGTGGGCTCACCAGAACGCGCGCTGCG 756  
QY 536 AGCTCGGACAATTTGGCAATTAGGGTTAAATTTGTTGCTCTCTTTTGGGGCTTCTACGCGTT 595  
DB 757 AGCTGCGCGCAGCGGGTCCGGGTCAACTGCTCTCGCCCTTCGGCGCTCGCCAGCCCA 816  
QY 596 TAGGCAGAA 605  
DB 817 TGCTCATCAA 826

RESULT 4  
US-09-504-358-13  
Sequence 13, Application US/09504358  
Patent No. 6365376  
GENERAL INFORMATION:  
APPLICANT: Bizostowicz, Patricia C.  
TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIA  
FILE REFERENCE: BC1001 US NA  
CURRENT APPLICATION NUMBER: US/09/504,358  
CURRENT FILING DATE: 2000-02-15  
EARLIER APPLICATION NUMBER: 60/120,702  
EARLIER FILING DATE: 1999-February-19  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 13  
LENGTH: 771  
TYPE: DNA  
ORGANISM: Brevibacterium sp HCU  
US-09-504-358-13

Query Match 8.6%; Score 70.2; DB 4; Length 771;  
Best Local Similarity 46.2%; Pred. No. 1.1e-12;  
Matches 348; Conservative 0; Mismatches 393; Indels 12; Gaps 3;

QY 29 GGCTAGAACGAAAGTTGCCCTTATAACAGAGGAGGCCAGTGGAAATGGAGAAACCCACAG 88  
DB 8 GACTCGCGGAAAGTAGCAGTCATTACTTGGGGGCCCGCAGGCATGGGGCGCATACAGT 67  
QY 89 CAAAACCTCTTCCCAACATGGAGCCAAAGTTGCCATTGCTGATGTCCAAAGATGAATTAG 148  
DB 68 CTGAACCTGTATGCGAGTGGAGGTGCAACAGTAGAGCGGTAGTAGATGTCATGAACAGAAG 127  
QY 149 GTCACTCACTGTTGCGAGGCCATTGGCACTTCCAA-----TTCCACCTACATCCACTGTG 202  
DB 128 GCGCTGCCACTGCCGATGCGATAAGGGCCAGCGGGGTTCGCAACTATTGGAATTTG 187  
QY 203 ATGTTACTAATGAAGCGGTGTAAATAATGCGGTGGACACACAGTTTCAACCTATGGAA 262  
DB 188 AGTTTCTGACGAGTCTGAAGTTGAATAATGTCGTCCTCCGACATTCGCAAGAGATTCGGTG 247  
QY 263 AACTGACATATTGTTCAAGCAATGCAAGAAATTTCTGATCCCAACAGCGCCCGCATCATAG 322  
DB 248 CGATTAACTACTAGTGAACACGAGCGCTCACCGGTGCGGATTAACCAACTCAC---G 304  
QY 323 ACAACGAAAGACGACTTTGAACGCGTCTCAGTGTAATGTAACCGAGTTTCTCTAT 382  
DB 305 AGATCGACGAACGGGACCTGGACCTCGTACTGAGCGTCTGATGTAAGAGGATTTCTTCA 364  
QY 383 GCATGAAGCAGCAGCAGCGCTGTATGATTCAGCAGCAGTGCACACATAATTTCCACTG 442  
DB 365 TGACAAACACATGCACTCCCTTCTTAAACAGGCTGGCGGCGGAGCCATCGCAACTTCG 424  
QY 443 CTAGTTTAAAGCTCAACTATGCGGTGGTGGTCTTCTACATGCTTATGTTGTTCAAGCATG 502  
DB 425 CGTCTATCTATGGTCTGGTGGGTCCAGAGGCTTACCCCGTACCAGCAGCAGCAAGGTG 484  
QY 503 CTGTGTTAGCCCTTACTAGGAATCTGGCAGTGCAGCTCGGACAAATTTGGCATTAGGTTA 562  
DB 485 CGGTCTGTTCCCTTACCAACAGGACGCGGTGACTTACGCGCGTCAATATCCGAGTGA 544  
QY 563 ATTGTTGTCCTTTCGGGCTTCTACGGCTTAGCGAAGAAATTT---TTCAGGAGTTA 619  
DB 545 ATGCGGTAGCAGCCCGGAACCATTTTGACTCCACTTCAAGGAGCTCGGTTTCAAGGGCC 604  
QY 620 AAAATGAAGAAATTTGAGAAATGTAATAAACTTTCCGGGAAATTTGAAGGTCCTCAAAAT 679  
DB 605 CCGATGGCTTAGATGATATATAACTTATGGGTGCCAAGCATCCGCTTGTGCGGTAG 664  
QY 680 TTAATGTTGAGGATGTTGCCAATGCAAGCTCTTTATCTGGCTAGTGATGAGGCAAAATACG 739

Db 665 GAACCCCGAAGAAGTCGGCGCAGCAACATCTTTCTGGCATCCGAAGAAGCTTCGTTC 724  
Qy 740 TGAGTGGACAATCTGTTTCATTGATGGAGGT 772  
Db 725 TTACTGGCGCCGCTCCCTCCGCTTGACGGTGGAT 757  
RESULT 5  
US-09-954-314-13  
; Sequence 13: Application US/09954314  
; Patent No. 6465224  
; GENERAL INFORMATION:  
; APPLICANT: Rouviere, Pierre E.  
; TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIATES  
; FILE REFERENCE: BC1001 US NA  
; CURRENT APPLICATION NUMBER: US/09/954,314  
; PRIOR FILING DATE: 2001-09-17  
; PRIOR FILING DATE: 60/120,702  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 13  
; LENGTH: 771  
; TYPE: DNA  
; ORGANISM: Brevibacterium sp HCU  
US-09-954-314-13  
Query Match 8.6%; Score 70.2; DB 4; Length 771;  
Best Local Similarity 46.2%; Pred. No. 1.le-12;  
Matches 348; Conservative 0; Mismatches 393; Indels 12; Gaps 3;  
Qy 29 GGCTAGAGCAAAAGTGGCCCTTATAACAGGAGGAGCCAGTGGATGGAGAACCCACAG 88  
Db 8 GACTCGCGCAAGTAGAGCATATTACTGGGGCGCCGAGGATGGGGCGGCATACAGT 67  
Qy 89 CAAACTCTTCTCCCAACATGAGCCAAAGTTGCCATTGCTGATGTCGCAAGATGAATTAG 148  
Db 68 CTGAACGTGATGCGAGTGAAGGTGCACAAGTAGCGGTAGTAGATGTCAAATGAACAAGA 127  
Qy 149 GTCACCTCAGTTCGAGGCCATTGGCAGTCCAA-----TTCCACCTACATCCACTGTG 202  
Db 128 GCCGTGCCACTGCGCATGGGATAGAGGCCGAGCGGGGGTTCGCAACTATTGGAAATTGG 187  
Qy 203 ATGTTACTAATGAAGCGGTGTTAAAAATGCGTGACACACAGATTTCAACCTATATGAA 262  
Db 188 ACGTTCTGACGAGTCTGAAGTTGAATAGTCTGTCGACATTGCCAAGAGATTCGGTG 247  
Qy 263 AACTGGACATTATGTTCAACAATGAGGAATTTCTGATCCCAACAGCGCCGCGCATCATAG 322  
Db 248 CGATTACGTAAGTGAACAACGCGAGCGTCCCGGTGCGGATTAACCAACTCAC---G 304  
Qy 323 ACAACGAAAAGCAGACTTTGAACGCTTCTCAGTGAATGTAACCGGAGTTTCCTAT 382  
Db 305 AGATCGACGACGCGGACCTGGACCTGCTACTGAGCGTCTGTAAGAGGATATCTTCA 364  
Qy 383 GCATGAGCAGCAGCAGCTGTTATGATCCAGCAGGAGTGGCAACATAATTTCCACTG 442  
Db 365 TGCAAAACACTCGATCCCTTAAACAGCGTGGCGGCGGAGCCATCGTCAACTCG 424  
Qy 443 CTAGTTAAAGCTCAACTATGGTGGTGGTTCATCATGCTCTATGTTGGTTCAAAGCATG 502  
Db 425 CGTCTATGCTATGGTGGGTGGGTCGACAGGAGCTTACCCCTACACGCGCAAGAGGTG 484  
Qy 503 CTGFGTTAGCCCTTACTAGGAATCTGGCAGTGGAGCTCGGACAATTTGGCATTAGGGTTA 562  
Db 485 CGGTCTGTCCTTCAACAAACAGGACGGGTGACTTACGGACCGTCAAAATATCCGAGTGA 544  
Qy 563 ATTGTTTGTCTCTTCGGGCTTCTACGGGTTTAGGCAAGAAATTT---TTCAGGGATTA 619  
Db 545 ATGGGTAGCACCCGCAACCATTTTGAATTAACCTTTGACTCCACTAGTCAAGGAGCTCGGTTCAAGGGCC 604  
Qy 620 AAAATGAAGAAGATTGAGAAATTAATAAACTTTGCGGGAAATTTGAAGGTTCCAAAT 679

Db 605 CCGATGCTTAGATGGATATACTAAACTTATGGTGCCCAAGCATCCGCTTGGTGGGTAG 664  
Qy 680 TTATGTTGAGGATGTTGCCAATGACAGCTTTTATCTGCTAGTATGATGAGCAAAATACG 739  
Db 665 GAACCCCGAAGAAGTCGGCGCAGCAACATTTGTTTGCATCCGAAGAAGCTTCGTTC 724  
Qy 740 TGAGTGGACACAATCTGTTTCATTGATGGAGGT 772  
Db 725 TTACTGGCGCCGCTCCCTCCGCTTGACGGTGGAT 757  
RESULT 6  
US-09-504-358-15  
; Sequence 15: Application US/09504358  
; Patent No. 6365376  
; GENERAL INFORMATION:  
; APPLICANT: Rouviere, Pierre E.  
; TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIA  
; FILE REFERENCE: BC1001 US NA  
; CURRENT APPLICATION NUMBER: US/09/504,358  
; CURRENT FILING DATE: 2000-02-15  
; EARLIER APPLICATION NUMBER: 60/120,702  
; PRIOR FILING DATE: 1999-February-19  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 15  
; LENGTH: 10629  
; TYPE: DNA  
; ORGANISM: Brevibacterium sp HCU  
US-09-504-358-15  
Query Match 8.6%; Score 70.2; DB 4; Length 10629;  
Best Local Similarity 46.2%; Pred. No. 5.le-12;  
Matches 348; Conservative 0; Mismatches 393; Indels 12; Gaps 3;  
Qy 29 GGCTAGAGCAAAAGTGGCCCTTATAACAGGAGGAGCCAGTGGATTTGGAGAACCCACAG 88  
Db 9139 GACTCGCGCAAGTAGAGCATATTACTGGGGCGCCGAGGATGGGGCGGCATACAGT 9198  
Qy 89 CAAACTCTTCTCCCAACATGAGCCAAAGTTGCCATTGCTGATGTCGCAAGATGAATTAG 148  
Db 9199 CTGAACGTGATGCGAGTGAAGGTGCACAAGTAGCGGTAGTAGATGTCATGAACAAGA 9258  
Qy 149 GTCACCTCAGTTCGAGGCCATTGGCAGTTCCAA-----TTCCACCTACATCCACTGTG 202  
Db 9259 GCCGTGCCACTGCGCATGGGATAGAGGCCGAGCGGGGTTCGCAACTATTGGAAATTGG 9318  
Qy 203 ATGTTACTAATGAAGCGGTGTTAAAAATGCGTGACACACAGATTTCAACCTATGAA 262  
Db 9319 ACGTTCTGACGAGTCTGAAGTTGAATAGTCTGTCGACATTGCCAAGAGATTCGGTG 9378  
Qy 263 AACTGGACATTATGTTCAACAATGAGGAATTTCTGATCCCAACAGCGCCGCGCATCATAG 322  
Db 9379 CGATTAAAGCTAGTAGTGAACAGCGAGCGGTCAACGGTGGCGGATTAACCAACTCAC---G 9435  
Qy 323 ACAACGAAAAGCAGACTTTGAACGCTTCTCAGTGAATGTAACCGGAGTTTTCCTAT 382  
Db 9436 AGATCGACGACGCGGACCTGGACCTGCTACTGAGCGTGTGATGTAAGAGGATATCTTCA 9495  
Qy 383 GCATGAGCAGCAGCAGCTGTTATGATCCAGCAGGAGTGGCAACATAATTTCCACTG 442  
Db 9496 TGCAAAACACTCGATCCCTTAAACAGGCTGGCGGCGGAGCCATCGTCAACTCG 9555  
Qy 443 CTAGTTAAAGCTCAACTATGGTGGTGGTTCCTCATGCTCTATGTTGGTTCAAAGCATG 502  
Db 9556 CGTCTATCTATGTTGGTGGGTGCGCAGGAGCTTACCCGCTACCAACGCCAAGAGGTG 9615  
Qy 503 CTGFGTTAGCCCTTACTAGGAATCTGGCAGTGGAGCTCGGACAATTTGGCATTAGGGTTA 562  
Db 9616 CGGTCTGTCCTTACCAACAGGAGCGGTGACTTACGGACCGTCAAAATATCCGAGTGA 9675



Db 247 GATGTTTGGTACACATGACGAGGAT-----TACCCAGAGTACTCTTATGCTCAAGATG 300  
Qy 328 GAAAAAGCAGACTTTGAACGCGTTCFACGTGTAATGTAAACCGGAGTTTTCCTATGATG 387  
Db 301 ACAGAACGACATTTTGAAGAGTGTCTCAAGGTCAATCTGACTGTGCTTTAAATATGACA 360  
Qy 388 AAGCAGCAGCAGCGTGTATGATTCAGCAGCAGTGGCAGCAACATAATTTCCACCTGCTAGT 447  
Db 361 CAATCAGTCTTGAACCGATGATGAAGCCAGAGAGGTGCTATCATTAATATGCTAGT 420  
Qy 448 TTAAGCTCAACTATGCTGTGTTCTTCCACATGCTATGTTGGTTCAAGACATGCTGTG 507  
Db 421 GTTCTGTTGATGGGATATGCTCAAGCTAATCTATGCTGCTTCAAGGCTGGCTTG 480  
Qy 508 TTAGCCCTTACTAGGATCTGCGAGTTCGAGCTCGGACAAATTTGSCATTAGGGTTAATGT 567  
Db 481 ATTGGCTTTACAGTCTGTGGCCGCGAGGTGCTAGTTCGGAATATACGAGTCAATGTG 540  
Qy 568 TTGTCCTCTTTCGGGCTTCTAGCGGCTTTAGCGCAAGAAATTTTCAGGGATTAAAGATGAA 627  
Db 541 ATTGCTCCAGGATGATGAGTCTGATATGACAGCTATCTTATCAGATAAGATTAAAGAA 600  
Qy 628 G 628  
Db 601 G 601

RESULT 9  
US-08-961-527-7/c  
; Sequence 7, Application US/08961527  
; Patent No. 6420135  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 19702 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-961-527-7

Query Match 7.3%; Score 59.6; DB 4; Length 19702;  
Best Local Similarity 46.1%; Pred No. 1.8e-08;  
Matches 280; Conservative 0; Mismatches 319; Indels 9; Gaps 2;

Qy 24 AAGAGGCTAGAGGAAAAAGTTGCCCTTTATAACAGAGGAGGCCAGTGAATTTGGAGAAC 83  
Db 16196 AATGAAGTAAAGACATAAAAAATATCTTTATACAGGTTGAGTCTGGAATTTGCTTTGC 16137  
Qy 84 CACAGCAAACTCTTCTCCCAACATGGAGCAAAAGTTGCCATTTGCTGATGTCCCAAGATGA 143  
Db 16136 CATGCCCAACAAGTTTCTCAAGCAGGAGCAACATTTGCTTTAAACAGTCTGTTGGGCAAT 16077  
Qy 144 ATTAGGTCACTCAGTTTCTGAG--GCCATTGGCACTTCCCAATTTCCACCTACATCCACTG 200  
Db 16076 CTCAGAAGAATTTGCTGCTGAGTTTTCACACTATGATCAAGGTGGTTCCCATTTTCAGG 16017  
Qy 201 TGATCTTTACTAATGAAGACGCTGTTAAAAATGCCGTGGACACACAGTTTCAACCTATGG 260  
Db 16016 AGATGATCAGATTTTTCAGACGCTAAGCGTATGATTTGATCAAGCTATTTGCAGAACTGGG 15957  
Qy 261 AAAACTGGACATTTGTTTCAGCAATGCAGAAATTTCTGATCCCAACAGGCCCGCATCAT 320  
Db 15956 TTCAGTATGATGTTTGGTCAACAATGCAGGGAT-----TACCCCAAGATACTCTTTATGCT 15903  
Qy 321 AGACAACGAAAAAGCAGACTTTGAACGCGTTCTCAGTGTAAATGTAAACGGAGTTTTCCT 380  
Db 15902 CAAGATGACAGAACAGATTTTGAAGAAAGTCTCAAGGTCAATCTGACTGGTGGCTTTAA 15843  
Qy 381 ATGCATGAAGCAGCAGCAGCTGTTATGATTTCCAGCAGCAGTGGCAACATAATTTCCAC 440  
Db 15842 TATGACACAATCAGTCTTGAACCGATGATGAAGCCAGAGAGGTGCTATCATTAATAT 15783  
Qy 441 TGCTAGTTTAAAGCTCAACTATGTTGGTGGTGTCTTTCACATGCTATTGTTGTTCAAGCA 500  
Db 15782 GTCTAGTCTTGTGTTGTTGATGGGAATATTTGTCACGCTAATTTATGCTGCTTAAAGC 15723  
Qy 501 TGCTGCTTAGCCCTTACTAGTAATCTGCGAGTTCGAGTTCGAGCAAAATTTGGCATTAGGT 560  
Db 15722 TGGCTTATTTGGCTTTTACCAAGTCTGTGGCACCGAGGTGCTAGTTCGGAATATACAGT 15663  
Qy 561 TAATGTTTGTCTCCTTTTCGGGCTTCTACGGCTTCTAGCGCTTTAGCAAGAAATTTTCAGGATTA 620  
Db 15662 CAATGATGATGCTCCAGGAATGATGAGTCTGATATGACCGCTATCTCTGCAGATAAGAT 15603  
Qy 621 AAATGAAG 628  
Db 15602 TAAGGAAG 15595

RESULT 10  
US-08-858-207A-8  
; Sequence 8, Application US/08858207A  
; Patent No. 6348328  
; GENERAL INFORMATION:  
; APPLICANT: Black, Michael  
; APPLICANT: Hodgson, John  
; APPLICANT: Knowles, David  
; APPLICANT: Nicholas, Richard  
; APPLICANT: Stodola, Robert  
; TITLE OF INVENTION: No. 6348328el Compounds  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/858,207A  
; FILING DATE: 09-MAY-1997







APPLICANT: Mitsuhashi, Kazuya  
TITLE OF INVENTION: NOVEL CARBONYL REDUCTASE, METHOD FOR PRODUCING SAID ENZYME, DNA  
TITLE OF INVENTION: ENCODING SAID ENZYME, AND METHOD FOR PRODUCING ALCOHOL USING SAID  
TITLE OF INVENTION: ENZYME  
FILE REFERENCE: 06501-050001  
CURRENT APPLICATION NUMBER: US/09/468,738A  
CURRENT FILING DATE: 1999-12-21  
PRIOR APPLICATION NUMBER: JP 1999-171160  
PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: JP 1998-363130  
PRIOR FILING DATE: 1998-12-21  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: Patentin Ver. 2.0, reformatted using WordPerfect 5.1  
SEQ ID NO 28  
LENGTH: 786  
TYPE: DNA  
ORGANISM: Bacillus subtilis  
US-09-468-738A-28

Query Match 6.2%; Score 50.4; DB 4; Length 786;  
Best Local Similarity 51.8%; Pred. No. 2.5e-06;  
Matches 145; Conservative 0; Mismatches 126; Indels 9; Gaps 1;  
QY 32 TAGAAGAAAGTTGCCCTTATACAGGAGGAGCCAGTGGAAATGGAGAAACACAGCAA 91  
DB 14 TAAAGGAAAGTCGTCGCTATTACAGGAGCTGCTTCAGGGCTCGGAAAGGCGATGGCCA 73  
QY 92 AACTCTTCTCCCAACATGGAGCCAAAGTTGCCAT-----TGCTGATGTCCAAGATG 142  
DB 74 TTCGCTTCGGCAAGGAGGAGCGCAAAAGTGTATTCACTATTATAGTAATAACAAGATC 133  
QY 143 AATTAGGTCACTCAGTTGTGCGAGGCCATTGGCACATTCCAATTCCACCTACATCCACTGTG 202  
DB 134 CGAACGAGGTAAAGAAGAGGTTCATCAAGCGGGCGGTGAAGCTGTTGCTGCCAAGGAG 193  
QY 203 ATGTTACTATGAAGAGGTGTTAAATAAGCCGTGGACACACAGTTTCAACCTATGGAA 262  
DB 194 ATGTCAGAAAGAGGAGATGTAAAAAATATCGTGAACGGCAATTAAAGGAGTTCGGCA 253  
QY 263 AACTGGACATTATGTCAGCAATGCAGGAATTTCTGATCC 302  
DB 254 CACTCGATATTATGATTAATAATGCCGGTCTTGAAAAATCC 293

Search completed: February 18, 2003, 17:47:38  
Job time : 114 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 17:21:59 ; Search time 114 Seconds  
(without alignments)  
3659.088 Million cell updates/sec

Title: US-09-673-918A-1

Perfect score: 819

Sequence: 1 atgcagcttcgaactgattc.....tgttccaatatccagattct 819

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	736	89.9	4505	9	US-09-944-160-23
2	150.2	18.3	774	9	Sequence 23, Appl
3	149	18.2	833	10	Sequence 945, Appl
4	123.6	15.1	391	10	Sequence 686, Appl
5	117.8	14.4	358	10	Sequence 2791, Appl
6	87.4	10.7	375	10	Sequence 98, Appl
7	78	9.5	6948	10	Sequence 3102, Appl
8	70.2	8.6	771	10	Sequence 4, Appl
9	70.2	8.6	10629	10	Sequence 15, Appl
10	62.4	7.6	774	9	Sequence 15, Appl
11	60.6	7.4	732	10	Sequence 55, Appl
12	60.6	7.4	732	10	Sequence 262, Appl
13	60.6	7.4	762	10	Sequence 9483, Appl
14	56.8	6.9	805	10	Sequence 7745, Appl
15	54.8	6.7	762	10	Sequence 797, Appl
16	52.8	6.4	2774	10	Sequence 2259, Appl
17	51	6.2	735	10	Sequence 5, Appl
18	50.4	6.2	786	10	Sequence 6029, Appl
19	49.2	6.0	1716	10	Sequence 28, Appl
					Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-09-944-160-23  
; Sequence 23, Application US/09944160  
; Patent No. US20020174452A1  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, NO. US20020174452A1man  
; APPLICANT: Davin, Laurence  
; APPLICANT: Huang, Ning  
; TITLE OF INVENTION: Monocot Seeds with Increased Lignan  
; TITLE OF INVENTION: Content  
; FILE REFERENCE: WSUR17983  
; CURRENT APPLICATION NUMBER: US/09/944,160  
; CURRENT FILING DATE: 2001-08-30  
; PRIOR APPLICATION NUMBER: US 60/230,632  
; PRIOR FILING DATE: 2000-09-07  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 23  
; LENGTH: 4505  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: plasmid pAPI249 encoding secoisolariciresinol  
; OTHER INFORMATION: dehydrogenase (dehydrogenase)  
US-09-944-160-23

Query Match 89.9%; Score 736; DB 9; Length 4505;  
Best Local Similarity 93.9%; Pred. No. 4.2e-204;  
Matches 766; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY	4	CAGCTTCGAACTGCATTCCGCAAGAGGCTAGAGGAAAGTTGCCCTTATAACAGGAGGA	63
Db	965	CAGCTTCGAACTGCATTCCGCAAGAGGCTAGAGGAAAGTTGCCCTTATAACAGGAGGA	1024
QY	64	GCAGTGGAAATTGGAGAAACACACGACAAACTCTTCTCCCAACATGGAGCAAAGTTGCC	123
Db	1025	GCCAGTGGAGTTGGAGAACTCACGACAAACTCTTCTCCCAACATGGAGCAAAGTTGCC	1084
QY	124	ATTGCTGATGTCACAGATGAATAGTCTACTGCTGAGGCCATTGGCACTTCCCAAT	183
Db	1085	ATTGCTGATGTCACAGATGAATAGTCTACTGCTGAGGCCATTGGCACTTCCCAAT	1144

QY 184 TCCACCTACATCCACTGTGATGTTACTAATGAAGAGCGTGTATAAAATGCCGTGGACAAC 243  
DB 1145 TCCACCTACATCCACTGCGATGTTACTAATGAAGAGCGTGTATAAAATGCCGTGGACAAC 1204  
QY 244 ACAGTTTCAACCTATGGAACCTGGACATATGTTTCAGCAATGCGAGGAATTTCTGATCCC 303  
DB 1205 ACAGTTTCAACCTATGGAACCTGGACATATGTTTCAGCAATGCGAGGAATTTCTGATCCC 1264  
QY 304 AACAGGCCCCCATCATAGACAACGAAAGAGAGACTTTTGAACGGTGTCTCAGTGTAAAT 363  
DB 1265 TACAAGCCCCGGGTATAGACAACGAAAGAGAGACTTTTGAACGGTGTCTCAGTGTAAAT 1324  
QY 364 GTAACCGAGTTTTCCTATGATGATGAGCGAGCGAGCTGTATGATTCAGAGCGAGCT 423  
DB 1325 GTAACCGAGTTTTCCTATGATGATGAGCGAGCGAGCTGTATGATTCAGAGCGAGCT 1384  
QY 424 GGCAACATAATTTCCACTGCTAGTTTAAGCTCAACTATGGGTGGTGTCTTTCACATGCC 483  
DB 1385 GGCTGCATAATTTTCCTGCTAGTTTAAGCTCAACTATGGGTGGTGTCTTTCACATGCC 1444  
QY 484 TATTGTGGTTCAAGCATGCTGTGTAGCCCTTACTAGAAATCTGGCAGTCCGATCGGA 543  
DB 1445 TATTGTGGTTCAAGCATGCTGTGTAGCCCTTACTAGAAATCTGGCAGTCCGATCGGA 1504  
QY 544 CAATTTGGCATATAGGTTAAATTTTGTCTCCTTTTCGGGCTTTCCTACGCGTTTATAGCAAG 603  
DB 1505 CAATTTGGCATATAGGTTAAATTTTGTCTCCTTTTCGGGCTTTCCTACGCGTTTATAGCAAG 1564  
QY 604 AAATTTTCAGGGATTAAATAATGAAGAATTTTGAATGTAAATAAATTTTCGGGAAAT 663  
DB 1565 AAATTTTCAGGGATTAAATAATGAAGAATTTTGAATGTAAATAAATTTTCGGGAAAT 1624  
QY 664 TTGAAGGTCCTCAAAATTAATTTTGAAGATTTTGAAGATTTTGAAGATTTTGAAGATTT 723  
DB 1625 CTGAAGGTCCTCAAAATTAATTTTGAAGATTTTGAAGATTTTGAAGATTTTGAAGATTT 1684  
QY 724 GATGAGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 783  
DB 1685 GATGAGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1744  
QY 784 AATCTGTAAATCAAGTTTCCAAATTCAGATTCCT 819  
DB 1745 AATCTGTAAATCAAGTTTCCAAATTCAGATTCCT 1780

## RESULT 2

US-09-938-842A-945  
; Sequence 945, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SAME, AND METHODS OF USE  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 945  
; LENGTH: 774  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-945

Query Match

18.3%; Score 150.2; DB 9; Length 774;

Best Local Similarity 52.7%; Pred. No. 6e-34;  
Matches 401; Conservative 0; Mismatches 348; Indels 12; Gaps 3;  
QY 28 AGGTTAGAAGAAAAGTTGCCCTTATAACAGGAGGAGCAGTGGAAATTTGGAGAAACCAACA 87  
DB 13 AGATTTGATGCAAAATCGCAATAATAACAGGCGGAGCTAGCGGATTTGGAGCCGAAGCG 72  
QY 88 GCAAAACTCTTCTCCCAACATGAGGCAAAAGTTGCCATTTGCTGATGTGCAAGATGAATTA 147  
DB 73 GTTAGGCTGTTCACGGACCCAGGAGCTAAGTGTGTCATCTGTTGACTTTCAAGAAGAACTT 132  
QY 148 GGTCACTCAGTGTTCGAGGCCATTGGCACTTCCAAATTCACCTACATCATCATGATGATTT 207  
DB 133 GGTCAAAAGTTTGGCGTTTCTGTCGGAAACAAAGCAAGTTTTACCGTTTCCGATGTT 192  
QY 208 ACTAATGAAGCGGTGTAAATAATGCGGTGGACAACAGTTTTCACACTATGGAAGAACTG 267  
DB 193 ACAACGAAAAGAGGTAGAAACGCCGTTAAGTTTACCGTCAAAAGATAGCGGAAGCTT 252  
QY 268 GACATTTATGTTACGAATGAGGAATTTCTGATCCCAACAGCCCGCATCATATAGACAAC 327  
DB 253 GACGTTCTCTTTAGTAACGCCGCGCTTATGGAAC---AGCCGGGAAGCTTTCTCGACTTG 309  
QY 328 GAAAAAGCAGACTTTTGAACGCTTCTCAGTGTAAATGTAAACCGGAGTTTTCCTATGCAATG 387  
DB 310 AATCTGGAACAGTTTGAACGCAACCATGGCGGTCAACGTTTCGTTGTCGCGTGCCTTTATC 369  
QY 388 AAGCAGCAGCAGCTGTATTGATTCCAGCAGCA---GTGGCAACATAAATTTCCACTGCT 444  
DB 370 AACCAGCTGCACGACCATGTTGGAGAAAGCAGCGGTGCTCAATCGTATGTAGAGAC 429  
QY 445 AGTTAAGCTCAACTATGGTGGTGGTGTCTTTCACATGCTATTTGTTTCAAGCATGCT 504  
DB 430 AGCGTCGGTCCGAGATCGGTGGTCCAGGACCTACCGCTACACGCGCTTAGCGCCCT 489  
QY 505 GTGTTAGCCCTTACTAGGAATCTGGCAGCTCGGACAAATTTGGCAATTTAGGTTAAT 564  
DB 490 CTCTCGGGCTGGTAAATCGGCTTGTGGCGGCTAGGGAAGTACGGGATTAGAGTCAAC 549  
QY 565 TGTGTTGCTCTTTTCGGGCTTCTACGGCTTTTAGGCAAGAAATTTTCAGGGGATTTAAAT 624  
DB 550 GCGCTTGCAACCATACGCGTGGCGGAGCGGATA-----AATAGCCGTGACGAGGAACG 603  
QY 625 GAAGAAGAAATTTGAGAATGTAAATAAATTTTCGGGGAATTTGAAAGCTCCAAAATTTAAT 684  
DB 604 GTGAGGATGGTGGAGGAATATAGCGCAGCCAGCGGATTTCTCAAAAGGTGTGTTGCTTAAG 663  
QY 685 GTTGAGGATGTTGCCAATGCACTCTTTATCTGGCTAGTGTAGGCAAAATACGTGAGT 744  
DB 664 GCTCGCCATGTGCGAGAGCGGCTTTGTTTCTGGCTTCGGATGATTCGGCTTACGTTAGC 723  
QY 745 GGACACAATCTGTTCAATTTGATGAGGCTTCAGCGCTCGAA 785  
DB 724 GGTCAATCTGCTGTGTGAGCGGGGTTATAGCGTGTAA 764

## RESULT 3

US-09-770-445-686/c  
; Sequence 686, Application US/09770445  
; Patent No. US20020023281A1  
; GENERAL INFORMATION:  
; APPLICANT: Gorlach, Jorn  
; APPLICANT: An, Yong-Qiang  
; APPLICANT: Hamilton, Carol M.  
; APPLICANT: Price, Jennifer L.  
; APPLICANT: Raines, Tracy M.  
; APPLICANT: Yu, Yang  
; APPLICANT: Rameaka, Joshua G.  
; APPLICANT: Page, Amy  
; APPLICANT: Matthew, Abraham V.  
; APPLICANT: Ledford, Brooke L.  
; APPLICANT: Woessner, Jeffrey P.  
; APPLICANT: Haas, William David

APPLICANT: Garcia, Carlos A.  
APPLICANT: Krieker, Maja  
APPLICANT: Slader, Ted  
APPLICANT: Davis, Keith R.  
APPLICANT: Allen, Keith  
APPLICANT: Hoffman, Neil  
APPLICANT: Hurban, Patrick  
TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
FILE REFERENCE: 2023US (PARA-012PRV)  
CURRENT APPLICATION NUMBER: US/09/770,445  
CURRENT FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: US 60/178,472  
PRIOR FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 999  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 686  
LENGTH: 833  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)-(833)  
OTHER INFORMATION: n = A,T,C or G

US-09-770-445-686

Query Match 18.2%; Score 149; DB 10; Length 833;  
Best Local Similarity 52.2%; Pred. No. 1.4e-33;  
Matches 395; Conservative 0; Mismatches 350; Indels 12; Gaps 3;

QY 32 TAGAGGAAAAGTTGCCCTTATAACAGGAGGAGCCAGTGGAAATGGAGAAACACACGAA 91  
DB 833 TGGATGGCAAAATCGCAATATAACAGCGGAGTAGCGGGATTGGAGCGGAAGCGGTTA 774  
QY 92 AACTCTCTCCCAACATGAGGACCAAGTTGCCATTCCTGATCTCCAAAGATGAATAGGTC 151  
DB 773 GCGTGTTCAGGACACGAGGTAGGTGTCATCGTTCACTTCAAGAAAGAACTGGTGC 714  
QY 152 ACTCAGTGTTCGAGGCAATTTGCCACTTCCAAATTCACCTACATCCACTGTGATTTACTA 211  
DB 713 AAAAGTTGCCGTTCTGTCGGGAAAGACAAAGCANNNTTTACCGTTCCGATTTACAA 654  
QY 212 ATGACAGCGGTGTTAAATGCGGTGGACACACAGTTTCAACCTATGAGAAACATGGACA 271  
DB 653 ACAGAAAGAGGTNNAAACGCGTTAAGTTCAACGTCGAAAAGTACGGGAAGCTTGAGC 594  
QY 272 TTATGTTACGAATGCAGGAATTTCTGATCCCAACAGGCCCGCATCATAGACACGAA 331  
DB 593 TTCTCTTTNNACCGCGCGTTATGGAAC---AGCCGGGAAGCTTCTCGACTTTGAATC 537  
QY 332 AAGCAGACTTTGAACGCGTTCTCAGTGAATGTAAACCGGAGTTTTCCTATGCATGAAGC 391  
DB 536 TGGACAGTTTCAACCAACCATGCGGTCACAGTTCGTGTCGGCTGCGTTTATCAAC 477  
QY 392 AGCAGCAGCTGTTATGATTCAGCAGCA---GTGGCAACATAATTTCCACTGCTAGTT 448  
DB 476 AGCCTGCACAGCCATGTTGGAGAAAGGACGCGTGGTCAATCGTATGTACGACGAGG 417  
QY 449 TAAGCTCAACTATGGTGGTGTCTTCACATGCTTATTTGTTCAAGCATGCTGCT 508  
DB 416 TCGGTGCGAGATGCGTGTTCAGGACCTTCACGCGTACACGCGTCAAGCAGCTCTTC 357  
QY 509 TAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGACAAATTTGGCAATTAGGCTTAATTT 568  
DB 356 TCGGCTGTTTAAATCGGCTTGTGGCGGTAGGGAAGTACGGGATAGAGTCAACGGCG 297  
QY 569 TGTCTCTTTCGGGCTTCTACGGCTTTAGGCAAGAAATTTTCAGGATTAATAATGAAG 628  
DB 296 TTGCACCATACGCGGTGGCGACGGGATA-----AATAGCCGTGACGAGGAACGGTGA 243  
QY 629 AAGAAATTCAGAAATGTAATAACTTTGCGGGAATTTGAAAGGTCCAAATTTAATGTTG 688  
DB 242 GGATGGTGGAGGAATATAGCGCAGCCAGGGGATTTCTCAAGAGGTGTGGTGTAAAGGCTC 183

QY 689 AGGATGTTGCCAATGCAGCTCTTTATCTGGCTAGTATGAGGCAAAATACGTAGTGGAC 748  
DB 182 GCCATGTGGCAGAGCGGCTTTTCTGGCTTCGGATGATTCGGCTTACGTAGCGGTC 123  
QY 749 ACAATCTGTTCAATGATGAGGTTTCAGGCTGTCGAA 785  
DB 122 AGAATCTGGCTGTTCACGGGGGTTATAGCGTCTGTTAA 86

## RESULT 4

US-09-878-574-2791  
Sequence 2791, Application US/09878574  
Patent No. US20020110548A1  
GENERAL INFORMATION:  
APPLICANT: Byrium, Joseph R.  
APPLICANT: Thompson, Michael D.  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
FILE REFERENCE: 38-21(15401)B  
CURRENT APPLICATION NUMBER: US/09/878,574  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: 09/333,535  
PRIOR FILING DATE: 1999-06-14  
NUMBER OF SEQ ID NOS: 15775  
SEQ ID NO 2791  
LENGTH: 391  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)-(391)  
OTHER INFORMATION: unsure at all n locations  
OTHER INFORMATION: Clone ID: LIB3028-020-Q1-B1-D11  
US-09-878-574-2791

Query Match 15.1%; Score 123.6; DB 10; Length 391;  
Best Local Similarity 60.8%; Pred. No. 2.3e-26;  
Matches 226; Conservative 0; Mismatches 134; Indels 12; Gaps 1;

QY 432 AATTCCACTGCTAGTTTAAAGCTCAACTATGGTGGTGTCTTTCACATGCTTATGTGG 491  
DB 1 AATTAACACTGCTAGTGTTCGCTGGAACCTTTAGTGGAGGGGCTTCACATGCTTACACAAG 60  
QY 492 TTCAAGCATGCTGCTTAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGACAATTTGS 551  
DB 61 TTCAAGCAGCAGCACTAATTTGGACTGATGAAACACTCGCGTGGAGCTTGGACAGTTGG 120  
QY 552 CATTAGGGTTAATTTGTTCTCTCTTCGCGGCTTCTACGGCTTTTAGGCAAGAAATTTTC 611  
DB 121 TATTAGGGTAATTTGTTCTCTCTCTCTATGTTGTCACACCATTCGACTAAGAAATGTTT 180  
QY 612 AGGGATTAATAATGAAGAATTTGAGAAATGTAATAAACTTTGGGGGAAATTTGAAGG 671  
DB 181 CAATCTGATGAAGACCGAAATGGTGAGA-----TTTATTCACACTAAAGG 228  
QY 672 TCCAAATTTAATTTGAGGATGTTGCCCATGCAGCTCTTTATCTGCTAGTGTGATGAGC 731  
DB 229 TGTTCATCTGTGCCAAACGATGTGGCGGAAGCTGCTCTATATTTGGCAGGTGATGAGTC 288  
QY 732 AAAATACGTGAGTGGACACAATCTGTTCAATTTGATGGAGGTTTCAGGCTGTGCAATTTCTGT 791  
DB 289 AAAGTATGTTAGTGGTCACAATCTTCTGTAGTGGAGGTTTCACCAATCTAATCTAGG 348  
QY 792 AATCAAGTGT 803  
DB 349 ATTTCTGTGT 360

## RESULT 5

US-09-878-574-98  
Sequence 98, Application US/09878574

Db	44	CTCCACTCTCCAAAGAGTTAGATGGCAAAAGTAGCACTCATAAACCGGGAGCCAGTGGCA	103
QY	74	TTCGGAACACCAAGCAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGATG	133
Db	104	TCGGTGAAGCCACCGCAAGCTTTTCTTCGGCCACGGTGCCAAAGTTCGTATCGCCGACA	163
QY	134	TCCAAGATGAATTAGTTCACCTCAGTTGTTCGAGGCCATTGGCACATTCCCAATTTCCA	187
Db	164	TCCAAGATAAACCCTCGGACACTCCCTATGTCCTCAAGTCTCAATTCCTCCGACAAAAACA	223
QY	188	-----CCTACATCCACTGTGATGTTTACTAATGAAGACGGTGTAAAAAATCCG	235
Db	224	ACGACGACATTTCCCTATGTTCACTCGGACGTCACCAACGACAAAGAGCTCGAAACCGCG	283
QY	236	TGGACAACACAGTTTCACCTTATGAAAACATGGACATATGTTGAGCAATGCAGAAATTT	295
Db	284	TCAACGCTCGGCTTCGGACACGCGCAAGCTCGACCTCTCTTTCAGCAACCGCGGCA	343
QY	296	CTG 298	
Db	344	CGG 346	

RESULT 7  
 US-09-070-927A-4/c  
 ; Sequence 4, Application US/09070927A  
 ; Patent No. US20020120116A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Charles A. Kunsch  
 ; Patrick J. Dillon  
 ; Steven Barash  
 ; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides  
 ; NUMBER OF SEQUENCES: 982  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/070,927A  
 ; FILING DATE: 04-May-2000  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/046,655  
 ; FILING DATE: 1997-05-16  
 ; APPLICATION NUMBER: 60/044,031  
 ; FILING DATE: 1997-05-06  
 ; APPLICATION NUMBER: 60/066,009  
 ; FILING DATE: 1997-11-14  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Kenley K. Hoover  
 ; REGISTRATION NUMBER: 40,302  
 ; REFERENCE/DOCKET NUMBER: PB369  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 6948 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
 ; IS-09-070-927A-4

Query Match		9.5%;	Score 78;	DB 10;	Length 6948;
Best Local Similarity		47.2%;	Pred. No. 2e-12;		
Matches 274;		Conservative	0;	Mismatches 300;	Indels 6; Gaps 1;
QY	32	TAGAAGAAAAGTTGCCCTTTATACAGGAGGAGCCAGTGGAAATGGAGAAACACACAGCAA	91		
Db	3145	TATCAGAAAAGTTATGTATTATCATGGGCGCTTCCAGTGGCATCGGTGAGGCAACAGCCCC	3086		
QY	92	AACCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGATGTCCTCAAGATGAATTAGGTC	151		
Db	3085	GTTTACTTGCAGAAAAGGAGCGAAGTTAGTCATTGCAGCAGCTGCGACAAGACGTTTAA	3026		
QY	152	ACTCAGTTGTGAGGCCATTGGCCATTCCTCAATTCACCTACATCCACTGTGATGTTACTA	211		
Db	3025	TTGCTATAAAAAAGAACTTCTTGAAGCAACAGATTCTCGTGCACACAGCAGATGTAACGA	2966		
QY	212	ATGAAGACGGTGTAAAAATGCCGTGGACAACACAGTTTCCAACTATGGAAAACCTGGACA	271		
Db	2965	AAGAAGAGGAAGTTACGGGTGTAAATCAAGCTTACAATGGAATAATATGGACGAATTCATG	2906		
QY	272	TTATGTTCCAGCAATGCAGGAATTTCTGATCCCAACAGGCCGCCATCATAGACAACGAAA	331		
Db	2905	TTCTTTTAAACAATGCAGGA-----GTTATGCCAACGGCGCTCTTATTGAAGCACCAA	2852		
QY	332	AACGACACTTTGAACGGTCTCTCAGTGTAAATGTAAACCGGAGTTTTCCTATGCATGAAGC	391		
Db	2851	AGGGAAGATGGCGTCAAAATGCTAGATATTAATATATATGTTGGTGTTTAAATGGCATTCGG	2792		
QY	392	ACGACGACGCTGTATGATTCACAGCAGCGATGCGCAACATAATTTCCACTGCTAGTTTAA	451		
Db	2791	CAGTTCTTCCAATATGTTGTGAACAAAATTCAGGGCAGATTAATGCTACTGATTCCTGTGG	2732		
QY	452	GCTCAACTATGGTGGTGGTTCTTCACATGCCCTATGTGGTTCAAAAGCATGCTGTGTAG	511		
Db	2731	CAGGACATGTCGTTTATCCAGATTTCGGCTGCTACTGTGGAACAAGTTTTCGTTCGAG	2672		
QY	512	CCCTTACTAGGAATCTGGCAGTCGAGCTCGGACAATTTGGCATTAGGGTTAAATGTTGT	571		
Db	2671	CAATTATGGAAGGATTGCGCAACAAGAACAGAGAGATAACATTAATCAACGATTATTT	2612		
QY	572	CTCCTTTTCGGGCTTCTACGGCTTTAGGCAAGAAATTTTC	611		
Db	2611	CACCAGTGCTGTACAACAGAACTTTATCAACAATTTTC	2572		
RESULT 8					
US-09-954-314-13					
; Sequence 13, Application US/09954314					
; Patent No. US20020127666A1					
; GENERAL INFORMATION:					
; APPLICANT: Rouviere, Pierre E.					
; APPLICANT: Brzostowicz, Patricia C.					
; TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIATES					
; FILE REFERENCE: BC1001 US NA					
; CURRENT APPLICATION NUMBER: US/09/954,314					
; CURRENT FILING DATE: 2001-09-17					
; PRIOR APPLICATION NUMBER: 60/120,702					
; PRIOR FILING DATE: 1999-February-19					
; NUMBER OF SEQ ID NOS: 49					
; SOFTWARE: Microsoft Office 97					
; SEQ ID NO 13					
; LENGTH: 771					
; TYPE: DNA					
; ORGANISM: Brevibacterium sp HCU					
US-09-954-314-13					
Query Match		8.6%;	Score 70.2;	DB 10;	Length 771;
Best Local Similarity		46.2%;	Pred. No. 1.2e-10;		
Matches 348;		Conservative	0;	Mismatches 393;	Indels 12; Gaps 3;
QY	29	GGCTAGAGAAAAGTTGCCCTTATACAGGAGGAGCCAGTGGAAATGGAGAAACCCACAG	88		
Db	8	GACTCGCGGAAAAGTAGCAGTCATTACTTGGGGCGCCGACGATGGGGCCCATACAGT	67		

QY	89	CAAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGATGTCCAAGATGAATTAAG	148		
Db	68	CTGAACCTGTATGCGAGTGGAGGTGCACAAGTAGCGGTAGTAGATGTCATGAACAAGAAG	127		
QY	149	GTCACTCAGTTGTCGAGGCCATTGGCACATTCCAA-----TTCCACCTACATCCACTGTG	202		
Db	128	GCGTGCCACTGCGCATGGCATTAAGGCCACGCGCGGGTTGCAAACTATTTGAAAATTGG	187		
QY	203	ATGTTTACTATGAAGCGGTGTTAAANAATGCCGTGGACACACAGTTTCAACCTATGGAA	262		
Db	188	ACGTTTCTGACGAGTCTGAAGTTGAANAATAGTCGTCCTCCGACATTGCCAAGAGATTCCGGTG	247		
QY	263	AACCTGACATTTATTTTCAGCAATGSCAGGAATTTCTGATCCCAACAGCCCGCCGATCATAG	322		
Db	248	CGATTAACGCTACTAGTGAACAACGCGAGCGCTCACCGGTGCCGATAAACCAACTCAC---G	304		
QY	323	ACAACGAAGAAGCAGACTTTTGAACGGTCTCTCAGTGTAAATGTAAACGGAGTTTTCCTAT	382		
Db	305	AGATCAGCAACGAGCGGACCTGGACCTCGTACTGAGCGTCGATGTGAAGGAGTATTCCTCA	364		
QY	383	GCATGAAGCAGCAGCAGCGTGTATGATTCACGACGCGAGTGGCAACATAATTTCCACTG	442		
Db	365	TGACAAACACTGCATCCCTACTTTAAACAGGCTGGCGCGGAGCCATCGTCAACTTCG	424		
QY	443	CTAGTTTAAAGCTCAACTATGGTGGTGGTTCTTTCACATGCGCTATTGTGGTTCAAAGCATG	502		
Db	425	CGTCTATCTATGCTCTGGTGGGTGCGCAGGAGCTTACCCCGTACCACGCAACGAGGTG	484		
QY	503	CTGTGTTAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGACAATTTGGCATTAGGGTTA	562		
Db	485	CGTCTGTTGCGCTTACCAACAGCAGCGCGTGACTTACGGACCGTCAAAATATCCGAGTGA	544		
QY	563	ATTGTTTGTCTCTTTCGGGCTTCTACGGCTTTTAGSCAAGAATTT---TTCAGGGATTA	619		
Db	545	ATCGGTAGCAGCCGGAACCATTTTGACTCCACTAGTCAAGGAGCTCGGTTCAAGGGGCC	604		
QY	620	AAAATGAAGAAGAAATTTGAGAATGTAAATAACTTTGCGGGAATTTGAAAGGTCCAAAAT	679		
Db	605	CCGATGCTTAGATGGATATATACTAAACTTATGGTGCCCAAGCATCCGCTTGGTGGGTAG	664		
QY	680	TTAATGTTGAGGATGTTGCCAATGCGAGCTCTTTATCTCGGTAGTATGAGGCAAAATACG	739		
Db	665	GAACCCCGGAAGAAGTCGCGGCGAGCAACATTTGTTCTGGCATCCGGAAGAGCTTCGTTC	724		
QY	740	TGAGTGGACACAATCTGTTCAATTGATGGAGGT	772		
Db	725	TTACTGGCGCGTCTTCCCGTTGACGGTGGAT	757		

RESULT 9

US-09-954-314-15

; Sequence 15, Application US/09954314

; Patent No. US20020127666A1

; GENERAL INFORMATION:

; APPLICANT: Rouviere, Pierre E.

; APPLICANT: Brzostowicz, Patricia C.

; TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIA

; FILE REFERENCE: BC1001 US NA

; CURRENT APPLICATION NUMBER: US/09/954,314

; CURRENT FILING DATE: 2001-09-17

; PRIOR APPLICATION NUMBER: 60/120,702

; PRIOR FILING DATE: 1999-February-19

; NUMBER OF SEQ ID NOS: 49

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 15

; LENGTH: 10629

; TYPE: DNA

; ORGANISM: Brevibacterium sp HCU

US-09-954-314-15

Query Match

Best Local Similarity

8.6%;

Score 70.2;

DB 10;

Length 10629;

Matches 10629;

Pred. No. 4.6e-10;

Matches	348;	Conservative	0;	Mismatches	393;	Indels	12;	Gaps	3;
QY	29	GGCTAGAGGAAAAAGTTGCCCTTATAACAGGAGGAGGACGTGGAATTTGGAGAAACACAG	88						
Db	9139	GACTCGCGGAAAAGTAGCATCTAATTCTGGGGCGCGAGGACATGGGCGCATACAGT	9198						
QY	89	CAAAACTCTTCTCCCAACATGGAGCCAAAAGTTGCCAATTTGCTGATGCCAAGATGAATTAG	148						
Db	9199	CTGAACCTGTATCGGAGTGGAGGTGCACAAGTAGCGGTAGTAGATGTCAATGAACAAGAAG	9258						
QY	149	GTCACTCACTTTCTCGAGGCCAATTTGGCACTTCCAA-----TTCCACCTACATCCACTGTG	202						
Db	9259	GCGTGCCTCACTGCCGATCGGATAAGGCCAGGCGGGGTGCAAACTATTGGAATTTGG	9318						
QY	203	ATGTTACTAATCAAGACCGTGTAAATAATCCGTGGACACACAGTTTCAACCTATGGA	262						
Db	9319	ACGTTTCTGACAGCTCTGAAGTTGAAATAGTCTGCTCCGACATTTGCCAAGAGATTTCGGTG	9378						
QY	263	AACTGGACAATTTATTTACGAATTCGACGAATTTCTGATCCCAACAGCGCCGACATCATAG	322						
Db	9379	CGATTAAACGTACTAGTGAACAAACGACGCGCTACCGGTGCCGATAAACCACTCAC---G	9435						
QY	323	ACAACGAAAAAGCAGACTTTTGAACGCGTTCTCAGTGTAAATGTAAACGGAGTTTTCCTAT	382						
Db	9436	AGATCAGCAACGGACACTGACCTCGTACTGAGCGTGCATGTGAAAGGAGTATTCTTCA	9495						
QY	383	GCATGAAGCACGACGACGCTGTATGATTTCCAGCACGACTGGCAACAATTTTCCACTG	442						
Db	9496	TGACAAAACACTGCATCCCTACTTTTAAACAGGCTGGCGCGGACCATCTGCAACTCG	9555						
QY	443	CTAGTTTAAGCTCAACTATGGTGTGTGGTTCTTTCACATGCCTATTCTGGTTCAAAAGCATG	502						
Db	9556	CGTCTATCTATGGTCTGTGGGTGCGCAGGAGCTTACCCCGTACCACGACGCAAAAGTG	9615						
QY	503	CTGTGTTAGCCCTTACTAGGAATCTGSCAGTGCAGCTTCGGACAAATTTGSCAATTAGGGTTA	562						
Db	9616	CGETCGTTCGCTTACCAAACAGGACGGGTGACTACGGACCGCTCAATATCCGAGTGA	9675						
QY	563	ATTGTTTGTCTCCTTTCGGGCTTCTTAGCGCTTTAGGCAGAAATTT---TTCAGGGAATTA	619						
Db	9676	ATCGGTACACCCCGAACCAATTTTGACTCCACTAGTCAAGGAGCTCGGTTTCAAGGGGCC	9735						
QY	620	AAATCAGAGAAGATTTGAGAAATGTATAAATTTTGGGGGAATTTTGAAGTGCCAAAT	679						
Db	9736	CCGATGGCTTATGGATATATCTAAACTTATGGGTGCAAGCATCGCTTGTTCGGGTAG	9795						
QY	680	TTAATGTTGAGGATTTGCCAATCAGCTCTTATCTGGCTAGTGTATGAGGCAAAATACG	739						
Db	9796	GAACCCCCGAAAGTCGCGGCACAACTGTTTGGCATCCGAAGAAGCTTCGTTCA	9855						
QY	740	TGAGTGGACACAATCTGTTTCATTGATGGAGGT	772						
Db	9856	TTACTGGCGCGCTCTCTCCCGTTCCAGCTGAT	9888						

## RESULT 10

US-09-891-641-55  
; Sequence 55, Application US/09891641  
; Patent No. US20020155612A1

; GENERAL INFORMATION:

; APPLICANT: Ye, Rick

; APPLICANT: Bedzyk, Laura

; APPLICANT: Wang, T.

; TITLE OF INVENTION: NATURAL

FILE REFERENCE: CL1686 US NA

CURRENT APPLICATION NUMBER: US/O

; CURRENT FILING DATE: 20  
: NUMBER OF SEC ID NOS: 87

```

; NUMBER OF SEQ ID NOS:
: SOFTWARE: Microsoft 97

```

; SOFTWARE: MLC  
; SEQ ID NO 55

```

, SEQ ID NO 33
:   LENGTH: 774

```

TYPE: DNA

ORGANISM: *Bacillus subtilis*

```

US-09-891-641-55

Query Match          7.68; Score 62.4; DB 9; Length 774;
Best Local Similarity 46.9%; Pred. N.2.2e-08;
Matches 358; Conservative 0; Mismatches 37; Indels 30; Gaps

Qy 39 AAAAGTTCGCCCTTATAACAGAGGAGGAGCCAGTGGAATTGGAGAAACACACAGCAAAACTCTT 98
Db 9 ACAAGTCGCTTTGGTGCAGGGGCTCGCGGGAATTCAGATTCGAAATCGCAGAGAAAT 68
Qy 99 CTCCTCAACATGAGCAAAAGTTGCCATTGCTGATGTCCAAAGATGAATTAGTGCACTCAGT 158
Db 69 CGCCGGGAAGGTGCCAGCGTCATCGTTTCAGACCTCCGTCGGAAGCATGTGAAAAGC 128
Qy 159 TGTGCG-----AGGCGATTGGCACTTCCCAATTCACACTACATCCACTGCTGATGTTACTAA 212
Db 129 AGCTCCAAAGCTTGCAGAAAGAGGCTTTGACGGGGCGCCATTTCGTTATGATGTGACAAA 188
Qy 213 TGAAGACGGTTTAAATGCGGTGGNACACACAGTTTCACCTATGGAANAATCGNACAT 272
Db 189 GGAAGCGCAAGTTGCTGATACGGTGAACGTTCATCAAAAACAAATACGGCCGCTTGGATAT 248
Qy 273 TATGTTTCAGCAATGAGGAATTTCTGATCCCAACAGGGCCGCAATCATAGACAACGAAAA 332
Db 249 TCTGTGTGAACAATGCCGGTAT-----TCAGCACGTTCGCTCCGATTTGAAGAGTTTCCGAC 302
Qy 333 AGCAGACTTTGAACCGGTTCTCAGTGTAAATGTAACTGACCGGAGTTTTTCCATGTCATGAAGCA 392
Db 303 AGACACCTTTGAACAGCTGATCAAGSTGTCATGCTGACGGCTCCCTTTATTGCAATGAAGCA 362
Qy 393 CGACAGCACGTGTATGATTCCAGCACACGAGTGGCAACATAATTTCCACTGCTAGTTTAAAG 452
Db 363 TGTGTTTCCGATCATGAAAAAACAGCAGTTTGGCAGAAATCATTAATATTGCGTCTGTAA 422
Qy 453 CTCACATATGGGTGGTGTCTTCCATATGCCATTCCTATGTGGTTTCAAGAGCATGCTGTGTAGG 512
Db 423 TGCATTAGTGGCTTTGCAGGAAAATCCGCTTATATAGCGCAAGCAGCGCTCATTTGG 482
Qy 513 CPTTACTAGTAATCTGGCAGTCGAGCTCGGACAATTTGCGCATTAGGTTTAAT-----564
Db 483 ACTCAAAAGTAGGGGCTTGAAGCGCGCCCAACGSCATACAGTCAATCGCTCTG 542
Qy 565 ---TGTTTCTGTCCTTTTCGGCTCTCTACGGCTTTTAGCCAGAAATTTTCAGGATTA 621
Db 543 TCCGGTTTATGTCGATACCCAGCTTTGTACCAATCAGCTTACCGATCTATCGAAAACACTAG 602
Qy 622 AA-----TGAAGAAGAAATTTGAGATGTAAATAACTTTGGGGAATTTGAAAGGTCG 674
Db 603 AAATGTCCCTTACGACTCTGTACTGTGAACAAGTCATTTTTCGCTTGTGGCCGCAAAAGCG 662
Qy 675 AAAATTTAATGTGTAGGATGTGTGCCATGCAGCTCTTTATCTGGCTAGTGTATGAGGCAAA 734
Db 663 ACTGCTTTCCGTCGAAGGAATTCGGGATATGCCGTGTTTTTGGCAAGCAGGAAGCGAA 722
Qy 735 ATACGTGAGTGACACAACTCTGTTTCATTTGATGAGGGTTTCAGCG 778
Db 723 GGCGCTCACTGGCAGGCTGTCGTCCTTTGATGGGGCTACACCG 766

```

RESULT 11

US-09-815-242-9262

; Sequence 9262, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.  
APPLICANT: Cxxx Cxxx T

APPLICANT: Yamamoto Robert T

APPLICANT: YAMAMOTO, ROBERT  
APPLICANT: XU, H. Howard

**: TITLE OF INVENTION:** IDENTIFICATION OF THE INVENTOR



```

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/915.242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/205,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9262
; LENGTH: 732
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(732)
US-09-815-242-9262

```

Query Match 7.4%; Score 60.6; DB 10; Length 732;  
Best Local Similarity 46.3%; Pred. No. 7.1e-08;  
Matches 278; Conservative 0; Mismatches 314; Indels 9; Gaps 2;

	Qy	31	CTAGAAGGAAAGTTCGCCCTTTATAACAGGAGGACCACAGTGGAATTCGGAGAAACCCACAGCA	90
	Dd	7	CTAGAACAATAAATATCTTTTTATTACAGGTTCGAGTCGTGGAATTCGTTCGCATCGCC	66
	Qy	91	AAACTCTTCCTCCCAACATGGAGCCAAGTTGCCATTGCTGATGTCCAAATGAATTAAGT	150
	Dd	67	CACAAAGTTTGCTCAGCAGAGGCCAACATTCCTTTAAACAGTCGTGGGCAATCTCAGAA	126
	Qy	151	CACTCAGTTGTCGAG--GCCATTGGCACATCCAAATCCACCTPACATCCACCTGTGTAT	207
	Dd	127	GAAATGTGCTCGCTGAGTTTCAAACTATGGTATCAAGGTGGTTCCTCATTTCCAGGAGATGA	186
	Qy	208	ACTAATCAAGACGGTGTAAAAATCGCGTGGCAACACAGATTTCAACCTATTGGAAAACATG	267
	Dd	187	TCAGATTTTTCAGACGCTAAGCGTATGATGTACAGCTATTTGCAGAACTGGGTTCAGTA	246
	Qy	268	GACATATATGTTTCAGCAATGCAGAAATTTCTGATCCCCAACAGGCCCGCATCATAGACAAC	327
	Dd	247	GATGTTTGTGCTAACATGCAGGGAT-----TACCCAAGATACTCTTATGCTCAAGATG	300
	Qy	328	GAAGAACGACATTTGAAACGGTTCFCAGTGTAAATGTAAACGGAGTTCCTTCTATGCAATG	387
	Dd	301	ACAGAAGCAGATTTTGAANAAGTGCTCAAGGTCAATCTGACTGGTGCCTTTAATATGACA	360
	Qy	388	AAGCAGCASCACGTGTTATGATTCAGCAGCAGTCGGCAACATATTTCCACCTGCTAGT	447
	Dd	361	CAATCAGCTTTGAACCGATGTGAAGCAGAGAGAGGTGCTATCATTAATATGCTTAGT	420
	Qy	448	TTAAGCTCAACTATGGTGGTGGTCTTTCACATGCCATTTGTTGGTTTCAAGGATGCTGTG	507
	Dd	421	GTTCCTGGTTGATGGGAATATTTGTCAGCTTAACATATGCTGCTTCTAAGGCTGGCTTG	480
	Qy	508	TTAGCCCTTTACTAGGAATCTGGCAGTCGAGCTCGGCAATTTTGGCATTTAGGCTTAATTCGT	567
	Dd	481	ATTGGCTTTTACCAAGTCTGTGGCAGCGAGGTGGCTAGTCGGAATATACGAGTCAATGTG	540
	Qy	568	TGTCTCCCTTTCCGGCTTCTACGGCTTTAGGCAAGAAATTTTCAGGGATTTAAAATGAA	627
	Dd	541	ATTGCTCCAGGAATGATGTAGTCTGATATGACAGCTATCTTATCAGATGAAGATTAAAGAA	600

```

QY      528 G 628
      |
Db      601 G 601

RESULT 12
US-09-815-242-9483
; Sequence 9483, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9483
; LENGTH: 732
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(732)
US-09-815-242-9483

```

Query Match	7.4%	Score	60.6;	DB	10;	Length	732;
Best Local Similarity	46.3%;	Pred.	No. 7.le-08;				
Matches	278;	Conservative	0;	Mismatches	314;	Indels	9; Gaps
2;							
QY	31	CTAGAAGAAAGTTGCCCTTATACAGGAGGAGCCAGTCGGAATTGGAGAACCACACAGA	90				
Db							
Db	7	CTAGAACAATAAATAATCTTTATTACAGGTCGACTCGTGGAAATGGTCTTGCCCATCGCC	66				
QY	91	AACCTCTTCTCCCACATGGAGCAGCAAAATGTGCATGTGCCAAGATGAATTAGGT	150				
Db							
Db	67	CACAAGTTTGCTCAAGCAGGAGCCAACATTGCTCTTAACACAGTCGTGGGCAATCTCAGAA	126				
QY	151	CACTCAGTTGTCGAG----GCCATTGGCACATTCCAAATTCACCTACATCCACTGTGATGTT	207				
Db							
Db	127	GAAITGTCGTGAGTTTTCAAACATATGGTATCAAGGTGGTTCCCAATTCAGGAGATGTA	186				
QY	208	ACTAATCAAGACGGTGTTAAAAATGCCGTGCACAAACACACAGTTTCAACCTATGGAAAAC	267				
Db							
Db	187	TCAGATTTTCAGACGGTAAAGCGTATGATTGATCAAGCTATTGCAGAACTGGGTTCAGTA	246				
QY	268	GACATTAATGTTACGAATGCAGGAATTTCTGATCCCCAACAGGCCCGGCATCATAGACAAC	327				
Db							
Db	247	GAUCTTTTGGTCACAATGAGGGAT-----CACCCAAGATACTCTTATGCTTTAAGATG	300				
QY	328	GA AAAAGCAGACACTTGAACCGCTTCTCAGTGTAATGTAAACCGGAGTTTTTCCCTATGCATG	387				



```

; LENGTH: 805
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-797

Query Match
Best Local Similarity 6.9%; Score 56.8; DB 10; Length 805;
Matches 252; Conservative 0; Mismatches 267; Indels 15; Gaps 2;

QY 269 ACATTATGTTTCAGCAATGTCAGCAATTTCTGATCCCAACAGGCCCGCCGATCATAGACAACG 328
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 ACATCTAGTTAACTACGTAGTGAATCTCAATCCCAATTAACCTACCATCGCCCAACACTC 343

QY 329 AAAAAAGACAGCTTTGACCGGTTCTCAGTGTAAATGTAACCGGATTTTCCCTATGCATGA 388
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 344 CCATTGAAGAATTCGATCGCATCTTCAAGGTGAACACAAGAGGATCATCTTATGCTGTA 403

QY 389 AGCACCAGCAGCGTGTATGATTCACGACGAGCGAGTGGCAACATAATTTCCACTGCTAGTT 448
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 404 AGAAGCAGCAAAAAGGCTAAACGCTGAGGCGGTGGTAGGATTATACCTAACGCTCGT 463

QY 449 TAAGCTCAACTATGGTGTGGTCTTTCATCGCTATTGTGTTCAAAAGCATGCTGTGT 508
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 464 CGTTACCGAGGGTTAAATCCCGGSCAGGAGCTTATACAGCATCAAGGCGAGCTGTG 523

QY 509 TAGCCCTTACTAGGAATCTGGCAGTCTGAGTCTGGACAATTTGGCATTAAGGTTAATTGTT 568
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 524 AAGCAATGGTGAAGATCTTGTCTAAGGAATTAAGGTTTAGGCATCACTGCAAACTGTG 583

QY 569 TGTCTCTTTCGGGCTTCTACGGCTTTAGGCAAGAAATTTTCAGGGATTAATAATGAAG 628
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 584 TATCTCCAGGCGCTGTGGGACGG-----AGATGTTTTTTGACGGGAAGAGTGAAG 634

QY 629 AAGAAATTTGAGATGTAAATAAATTTTCGGGAAATTTGAAAGTCCAAATTTAATGTTG 688
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 635 AGACGGTGAATATCA-----TTGAGAGGAGTCCCTTTGGTAGGCTGGTGAGACTA 688

QY 689 AGGATGTTGCCATGACGCTCTTTATCTGGCTAGTAGTAGGCAAAATACGTAGTGGAC 748
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 689 AAGATATTCTCTGTTGTTGTTTCTTAGCTAGTAGTGTGAGTGTGATGATCAATGGAC 748

QY 749 ACAATCTGTTATGATGAGGTTTCAGGCTCGCAATTCGTGAATCAAAAGTGT 802
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 749 AAGTTATTGTTGTAATGAGTGCATTCCTCAAAATGAATATGAGTTATGTTGTTT 802
```

```

RESULT 15
US-09-974-300-2259
; Sequence 2259, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2259
; LENGTH: 762
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2259

Query Match
Best Local Similarity 6.7%; Score 54.8; DB 10; Length 762;
Matches 186; Conservative 0; Mismatches 192; Indels 6; Gaps 1;

QY 180 CAATTCACCTACATCCACTGTGATGTTTACTATGAAGACGGTGTAAAAATGCCGTGGA 239
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151 CAAATCGCTGTGGATTCAGCGGATTTTACAAAGGAAGCCGATGTGAAAAAATGATCGG 210

QY 240 CAACACAGATTTTCAACCTATGGAATACTGGACATTTATGTTTCAGCAATGCGAGGAATTTCTGA 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 211 CAGGGTTCAGGGAGCAATTCATTAATAATCGATGTGATGGTGTACAGTTCGCGGAATCGCAA 270

QY 300 TCCCAACAGGCCCGCCGATCATGACACAACGAAAGACAGACTTTTGAACGCGTTCAGTGT 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 271 AAGCAATA-----AAATCAGGATTTTGAACITGAAGACTGGGAACGGACCATCGAGGT 324

QY 360 AAATGTAACCGGAGTTTTCTCTATGATGAAGCAGCAGCAGCTGTTTATGATTCAGCAGC 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 325 CAATTTGACCGGGTATTTTCTTTGCGCGAGGGAACGTCAAAAGGTCATGATTCAGCAGG 384

QY 420 CAGTGGCAACATAATTTCCACTGCTAGTTTAACTATGAGCTCAACTATGGGTGGTTCCTTACA 479
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 385 AAGCGGGTCAATCATCTTCAAAATCCAAATCGGGAAGTTCGGCAGCAAGCATATATAC 444

QY 480 TGCCTATTGTGTTCAAAAGCATGCTGTGTAGCCCTTACTAGGAATCTGSCAGTCGAGCT 539
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 445 AGCTTATTCAGCTTCTTAAATTTGGAGGGTTCGTCTCACGCAAAAGCCTCGCGTTGATCT 504

QY 540 CGGACAATTTGGCATTAGGGTTAA 563
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 505 TGCCGAACACAAACATTTCGGGTGAA 528
```

Search completed: February 18, 2003, 18:33:02  
Job time : 131 secs

**THIS PAGE BLANK (USPTO)**  
**THIS PAGE BLANK (USPTO)**